T H E U N I V E R S I T Y O F A L B E R T A

RELEASE FORM

NAME OF AUTHOR.....Cuthbert Horatio C. JOSEPH.....

TITLE OF THESIS.....Academic Self Concept and the Differential Influences of Immediate and Delayed Feedback on Level of Performance in Arithmetic Skill.....

DEGREE FOR WHICH THESIS WAS PRESENTED.....Master of Education.....

YEAR THIS DEGREE GRANTED.....1981.....

THE UNIVERSITY OF ALBERTA

Academic Self Concept and the Differential Influences of Immediate and Delayed Feedback on Level of Performance in Arithmetic Skill

by

© Cuthbert H. C. Joseph

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF

MASTER IN EDUCATION

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

(SPRING), (1981)

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled ACADEMIC SELF CONCEPT AND THE DIFFERENTIAL INFLUENCES OF IMMEDIATE AND DELAYED FEEDBACK ON LEVEL OF PERFORMANCE IN ARITHMETIC SKILL.....

submitted by CUTHBERT H. C. JOSEPH.....

in partial fulfilment of the requirements for the degree of

MASTER OF EDUCATION

ABSTRACT

This study examined the effects on performance in arithmetic of varying the time of giving feedback. Data on arithmetic performance was collected on three occasions, and performance was assessed for effectiveness at several levels of self concept of academic ability.

The sample consisted of 236 Grade 4 students in elementary schools of Edmonton. They were all at this grade level during the 1980-81 school year. The instruments used to measure performance in arithmetic were three forms of an achievement test, each composed of 26 four-option multiple choice items. A 3M Model 550 Test Scoring Computer was used to provide feedback on the students' performance. It allowed students to obtain their results promptly. Academic self concept was measured by the Student's Perception of Ability Scale (Boersma and Chapman, 1977; Boersma, Chapman and Maguire, 1979).

The achievement tests were designed to provide students with practice on skills they were taught previously, and were used as a mode of instruction. At each of the three occasions, the treatments were characterized by an attempt at one of the test forms, feedback on each item response, a second or subsequent attempt until students could do each exercise correctly, or until they were satisfied with their score. In one treatment, students were issued with knowledge of results (KR) immediately after they had completed each test. The other KR treatment group received feedback one day after completing each test. The control

group was not given any feedback until they had completed all three test forms, at which time they were made to know their scores on each of the three test forms. The student's preference for KR at either of these three times was examined in a post hoc analysis.

The results were all in the direction expected. As predicted, no difference was observed between the effectiveness of immediate post-test knowledge of results (IKR) and one-day delayed knowledge of results (DKR) ($p > .10$). Both KR treatments, however, were associated with highly significant increases in level of performance ($p < .001$). Self concept of ability in arithmetic was, in all instances, a very significant predictor of level of arithmetic performance. Self concept of ability in reading, on the other hand, indicated a significant pooled regression estimate only after KR was issued. This latter concomitant interacted very significantly with the effectiveness of the treatments, though it did not show a significant interaction on measures taken before the first issue of feedback. On all post KR measures, the reading self concept regression estimate was higher for the IKR treatment than for the DKR treatment, and showed that it was the better treatment for students of very high self concept of ability in reading. The regression estimates for self concept of ability in arithmetic indicated the same trend though the corresponding differences did not reach statistical significance. No difference was observed between the students' preferences for feedback given immediately after completing the test and one day after completing the test. However, when a third option (feedback one week after completing the test) was considered in the analysis, the findings indicated that students in this

population preferred shorter than longer delays of feedback.

The findings reported in this study support the view that delaying feedback for up to one day does not have identifiable deleterious effects on students' performance in arithmetic. They also provide some support for practices in Computer Assisted Instruction where students are frequently provided with feedback within a few seconds after they have completed a test. Findings with respect to the students' preferences, as well as the fact that students treated with immediate posttest KR showed greatest preference for this type of feedback, gave added prominence to innovations which favour immediate posttest KR. It was recommended that further research be done to investigate the interactions of academic self concept and feedback preference with the effectiveness of feedback.

ACKNOWLEDGEMENTS

The writer hereby makes known his recognition of indebtedness to others, and extends many thanks to especially those persons mentioned in this statement of acknowledgements.

A special gratitude is extended, firstly, to his supervisor, Dr. T. O. Maguire, who was ever available to listen, encourage, and guide whenever these types of reinforcement or promptings were appropriate. Similar thanks go to other members of the committee (Dr. F. Boersma, and Dr. D. Sawada).

Thanks are extended, also, to Dr. T. Blowers (Director of Research, Edmonton Public Schools) and to the various teachers of the classes involved in the study. These teachers were particularly approachable and willing to make adjustments necessary to properly accomodate the data collection.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I

THE PROBLEM

Background to the Problem

The teacher and other implementers of educational technology have long recognized positive effects of feedback on students' performance. Such effects have been consistently demonstrated in studies such as Buchwald (1967, 1969), Buchwald and Meager (1974), Sassenrath (1975), Surber and Anderson (1975), O'Neil, Rasor and Bartz (1979), and Bloom and Bourdon (1980). Irrespective of differences in the intervals of time employed between initial trial - feedback - and subsequent trial, these studies have all demonstrated that knowledge of results is superior to no feedback, where the target behaviour is increased level of performance.

Yet, in spite of the large accumulation of empirical evidence, explanations as to how feedback influences level of performance are unclear and contradictory. What we have may be best described as a large body of experiments which relate one or two dimensions of feedback to a given response, with little or no regard for the psychological processes activated by the feedback stimulus.

Several attempts at formulating a theoretical framework (Hammond and Summers, 1972; Kantowitz, 1974; Deci, 1975; Hackman and Oldham, 1975, 1976) lean for support on subjects in job oriented organizations, and consequently, prove too far removed or extended in scope to be

meaningfully applied to students in a classroom learning situation. For example, the three most recent of these works propose a motivational theory which views performance feedback as carrying out a motivational function by providing the proper environment to meet higher order needs through task accomplishment. By 'motivational function' it is not clear whether these authors speak of an incentive or a reinforcement effect of feedback. Besides, several studies and reviews of the relevant literature (Annett, 1969; Tait, Hartley and Anderson, 1973; Surber and Anderson, 1975; Ilgen, Fisher and Taylor, 1979) have questioned the incentive and reinforcement capacities of knowledge of results. These theories, then, tend to describe an undifferentiated feedback concept which neither explicates the specific characteristics of feedback nor considers the ramifications of the different variations of the construct on behaviour (Herold and Greller, 1977). A correction of the present state of affairs seems possible only through an appreciation of the complexity of the feedback stimulus and of the likelihood that it interacts with the psychological idiosyncrasies of the individual making the response.

Some studies have considered specific characteristics of the feedback stimulus. Of interest here are those which have attempted to differentiate between the effectiveness of immediate versus delayed knowledge of results. There is now a recognizably large body of evidence showing delayed feedback as being at least as effective as immediate feedback in improving students' ability to recall verbal material. Some frequently referred to studies in this area are Buchwald (1967, 1969), Sassenrath and Yonge (1969), Sturgis (1969, 1972), Kulhavy and Anderson

(1970, 1972), Buchwald and Meager (1974), Surber and Anderson (1975), and Peeck and Tillema (1979). Several of these works go as far as to suggest that delayed feedback is superior to immediate feedback, a finding which seemingly contradicts the position held by most conventional behaviouristic theories of learning (Hull, 1952; Spence, 1956; Skinner, 1954, 1974) that learning is best when some feedback or reinforcer immediately follows the correct response. In addition, experiments and reviews of the literature on delay of reinforcement with animals (Renners, 1964; Ainslie, 1974; D'Amato and Cox, 1976, for example) indicate that efficiency in learning and retention decreases with delay of reinforcement. In most instances, learning decrements reached significance after delays as short as eight seconds.

The research with human subjects is not without contradictions. Some studies (Phye and Baller, 1970; Spartz and Sassenrath, 1972; Kippel, 1974; Beck and Lindsey, 1979; and O'Neil, Rasor and Bartz, 1979) reported no difference between the effectiveness of immediate and delayed feedback, while a few earlier studies (Greenspoon and Foreman, 1956; Sassenrath, Yonge and Shrable, 1968; and White, 1968, for example) showed immediate knowledge of results as being superior to delayed knowledge of results.

One limitation of this area of research has to do with the form and content of the tests used. They take the form of comprehending a written passage, or of recalling nonesense syllables or words. Peeck and Tillema (1979) attempted to improve on this limitation by using three types of test items. However, they concentrated more on varying the level

of mental complexity of their test items than on shifting away from the verbal bias of instruments used in these studies. Items on arithmetic operations, for example, would have been such a shift.

Another limitation has to do with the confounding of time intervals with feedback delay. When groups of students are treated with feedback of varying delay intervals, the experiences of the treatment groups differ not only with respect to degree of feedback delay but also with respect to the interval of time between trials. The three time intervals to contend with may be portrayed as follows: t_1-- f_1, f_1-- t_2, and t_1-- t_2; where t stands for trial, f for feedback, and -- for the interval between. In the immediate feedback situation, t_1-- f_1 is zero. This cannot be so for the delayed feedback group. If f_1-- t_2 is to be held constant across treatment groups, this would result in varying time periods between t_1 and t_2. Time may be a crucial factor of learning. On Trial 2, the fact that students in the immediate feedback group have been more recently exposed to the test items than were students treated with delayed feedback may have contributed to the observed differences between the performance of these two treatment groups. As concluded by Bilodeau and Bilodeau (1966), not many of the early studies in this area of research have attempted to offset differences attributable to the confounding of feedback delay with time between trials. This inadequacy may be partly responsible for the inconsistencies in the findings reported.

When performance feedback is connected to disciplines the subject has been learning over an extensive period of time, the impact of feedback on a person's behaviour is likely to be a function of several variables, both experiential and psychological. Amongst the psychological variables that seem important is academic self-concept. Since academic self-concept has been seen to be related to achievement and incidents of learning disability (Boersma, Chapman and Battle, 1979), it seems on the surface that it should also be related to how a student makes use of feedback. No study of performance feedback on students has included a self-concept variable, nor has anyone attempted to use this factor to improve the experimental design. Greater precision and fewer contradictions may be the biproduct of such tight controls.

The Specific Problem of the Present Study

The present study was designed to examine differences in arithmetic skill level between groups of fourth grade students who were provided with either immediate or delayed knowledge of results from skill tests. The results were analysed using measures of self concept of academic ability as covariates, and a post hoc examination of the effects of the treatments on the students' preferences was undertaken also. The treatments were evaluated for effectiveness using measures taken at three times, which were an initial occasion, and one day and seven days after feedback was issued. The study differs from all previous ones in this area not only because it included a self-concept variable, but also in that it used a different criterion - performance in arithmetic - in place of the repeatedly used recall of verbal material.

The Research Questions

The following research questions formed the basis of the study:

1. Does feedback influence performance in arithmetic skill?

2. Is there a significant difference between the effectiveness of immediate and delayed knowledge of results in relation to arithmetic skill performance?

3. What is the nature of the interaction between self-concept of ability in arithmetic and the effectiveness of feedback?

4. Would students put through this experiment prefer to know their results immediately, one day, or one week after they have completed each test?

Definition of Terms Used in This Study

Feedback

As used in this study, 'feedback' was simply a report of each item response as correct or incorrect. The student's responses were scored by a portable electronic scorer. Correct responses remained unmarked, whereas incorrect ones had a short red dash beside them. The total number of correct responses constituted the individual's score.

Immediate Feedback

'Immediate feedback' was defined as knowledge of results given immediately after the test was completed. The abreviation 'IKR' refers to this type of feedback.

Delayed Feedback

The term ' delayed feedback' (DKR) was defined as knowledge of results issued one day after the students had completed each test.

Academic Self-Concept

This construct had to do with the subject's perception of his academic ability, and includes his manner of distinguishing himself from all others on the academic skills in question. The 'Student's Perception of Ability Scale' (Boersma and Chapman, 1977) was used to measure this construct. Among its subscales were 'Self Concept of Ability in Arithmetic' and 'Self Concept of Ability in Reading'. The measure of the student's self-concept of ability in arithmetic included his ratings of his ability to do specific aspects of the discipline: for example, to do subtraction, or to work with numbers in general. It taps also the feelings like unhappiness or pleasure that the student associates with the subject.

II

SURVEY OF THE RELATED LITERATURE

Feedback: Knowledge of Results

Introduction

The term 'feedback' has been frequently used to mean knowledge of results (KR). It includes information about the subject's behaviour, and comes to him from sources such as his own perceptions of his performance and the assessments of others in his social setting.

'Intrinsic KR' refers to knowledge of results which is normally present and is not subject to experimenter manipulation. Feedback which is supplied or manipulated by the experimenter is generally termed 'extrinsic KR'. Obviously, much of feedback in any situation is intrinsic, but when the experimenter incorporates an additional feedback loop the entire system may be referred to as extrinsic or augmented feedback. In making this distinction between intrinsic and extrinsic KR clear, it should be noted that most 'no-feedback' situations (used in studies with a control group) leave the intrinsic feedback loops intact. Studies reporting an ineffective feedback treatment, then, should be interpreted with much caution, since the finding of no significant difference between treatment and control groups may not mean that the entire system of feedback is ineffective. One is permitted to say merely that the particular manipulations introduced by the experimenter did not

enhance the criterion over and above any influence attributable to intrinsic aspects of the system of feedback.

The findings of several studies (Buss and Buss, 1956; Michael and Maccoby, 1961; and Parkinson, 1964, for example) where Pressey type and Crowder type tests (with high error rates) were used, leave little doubt that knowledge of results is likely to have a significant effect on several types of learning. It is generally accepted that repeated success or repeated failure has an enhancing or a detrimental effect on performance. However, there are doubts as to whether it is the reinforcing capacity of KR, the conjoint information, or its incentive or motivational attributes that influence performance.

Information Theory

The proposition that delayed KR is at least as effective as immediate KR leans for support on information processing theory. If it is the informative value of KR that makes a difference, then, except where time or memory is a crucial variable, whether the information in feedback is given immediately after the response or is delayed, should have no differential effect on responses made subsequent to its arrival. Therefore, it may be worthwhile to look at information processing theory against the posits of reinforcement theory.

At the initial stages of the development of a theory of information, Shanon and Weaver (1949) equated information with the reduction of

uncertainty. They reasoned that if stimuli are seen as messages, the information they convey is not so much a function of the stimuli themselves as of what other stimuli might have occurred. Thus the informative value of a specific piece of feedback depends on how many different kinds of knowledge could have been sent as a result of that response. On a four option multiple-choice item, therefore, 'right' may be more informative than 'wrong'. Whereas 'right' reduces all uncertainty, the result 'wrong' would rule out only one of the four possibilities.

Bilodeau (1952, 1953), Bilodeau and Rosenback (1953), and Annett (1969) produced evidence which suggest that error information is used strategically to locate the unseen target, or the right response, rather than as a mere negative reinforcer. In other words, wrong responses have the positive effect of directing rather than inhibiting the subject's next response. This finding is consistent with the view that it is the informative value of KR that has an effect on performance.

However, it may be just as naive to assume that negative responses do not detrimentally affect behaviour as it is to think that that is all they do. Such a conclusion would contradict the bulk of the research dealing with repeated failure or feedback sign on performance. Bearing both types of research in mind, a more reasonable assumption seems to be that the extent to which the individual is inhibited by a negative result depends at least in part on his perception of his chances of succeeding on a subsequequent trial.

Several tests are used to examine the information versus reinforcement hypothesis. All of these involve the systematic withholding of information. A frequently used technique is referred to as partial KR. Here, treatment groups are given different reward frequencies: for example, KR at the end of every trial, every other trial, or every fourth trial. Another is ill-defined as 'blank trials'. It is simply giving nothing where the subject expects to receive feedback information.

With respect to partial reinforcement, the standard study is that conducted by Bilodeau and Bilodeau (1958). The task involved pulling a lever through an arc of 33 degrees. KR included a measurement to the nearest degree of the arc pulled with the words 'too high' or ' too low', since the subject was not told the value of the correct arc. The study used four treatment groups: one was given KR after every trial; another received KR after every third trial; a third had KR after every fourth trial; and a fourth was treated with KR after every tenth trial. Subjects in each treatment group continued to practice at the task until they received ten trials with KR. The number of errors in the ten trials immediately following KR was found to be approximately equal for all treatment groups. The amount of learning accrued, therefore, seemed related to the absolute rather than the relative frequency of KR. No measures were made of performance subsequent to the complete removal of KR. A finding of no difference here would sharply contradict evidence from conditioning studies.

In a similar study, James and Rotter (1958) included measures of performance subsequent to the complete removal of KR. Their study involved two treatment groups. One was given KR after every trial, while the second had KR after every other trial. The results obtained supported Bilodeau's (1958) findings and were typical of partial reinforcement in conditioning studies. In addition, under the skill conditions, 100 per cent reinforcement was not significantly (though slightly) slower to extinguish than 50 per cent reinforcement.

The hypothesis that subjects given blanks where they expect information would not show learning increases is apparently contradicted by the research on concept learning. Experiments in this area usually leave out either 'right' or 'wrong' in a series of feedback information, or give full results for some items and nothing for others. From a review of studies such as Buss and Buss (1956), Bourne, Guy and Wadsworth (1967), and Moore and Halperin (1967), Annett (1969) concluded that a blank trial is not neutral but is more like 'right', and that the combinations 'right or wrong', 'nothing or wrong', and 'right or nothing' are not equivalent to 'nothing or right' (p. 150). At least in some instances, giving the subject no feedback does not prevent him from registering an evaluation of his performance. He may reason that silence is equivalent to 'right' since the teacher would have corrected him if he were wrong. The fact that 'I am not marked wrong' is information. Levine, Leitenberg and Richter (1964) examined this explanation closely and came up with the 'blank trials law' which stated their position that blank trials are equivalent to positive reinforcement.

Still in concept learning, Buss and Buss (1956) used geometric forms drawn in various colours and shapes on separate cards. Subjects were required to sort the cards by colour and then by shape independently of the other dimension. The results contradicted explanations posited by reinforcement theory. However, they offered a reinforcement explanation based on the questionable assumption that 'right' is only a weak positive reinforcer, whereas 'wrong' is strongly negatively reinforcing. The weight of evidence from other studies in this area make quite the opposite assumption that 'right' is a strong positive reinforcer and 'wrong' is neutral or slightly positively reinforcing (see Thorndike, 1933).

Reexamining Buss and Buss' (1956) claim, Bourne et al (1967) systematically varied the proportions of trials in which 'right', 'wrong' and no-KR were given, and showed that when the number of trials accompanied by KR were held constant, differences between 'right or wrong', 'nothing or wrong' and 'right or nothing' disappeared. These findings contradict a reinforcement based explanation but at the same time raise questions about the exclusiveness of any interpretation based soley on information theory, since it was the proportions rather than the absolute frequencies which were seen to be related to performance. However, whereas reinforcement explanations lead to questionable conclusions, it seems much easier to account for these results if KR is regarded as having primarily an informative value.

Immediate versus delayed feedback. The Law of Effect requires that reinforcement be close in time to the behaviour which is to be reinforced. Close temporal contiguity is necessary if the hypothesized automatic nature of reinforcement is to be upheld. If response strength is increased or maintained after an extended period of time between the reinforcer and the response, the operation of mediatory processes will have to be accredited with significance, and the law of effect phenomenon would loose much of its force as an automatic, inevitable and simple biological process.

Several studies have examined the relative efficacy of immediate versus delayed feedback. Immediate feedback is usually KR given as soon as possible after each response. With human beings, delays of less than eight seconds are insignificant (Peeck and Tillema, 1979). Generally, the term 'delayed feedback' is used to describe situations where KR is withheld for hours or days.

Studies prior to 1965 (Lorge and Thorndike, 1935; Alexander, 1951; Greenspoon and Foreman, 1956; Bilodeau and Bilodeau, 1958b; Bilodeau and Ryan, 1960; Becker, Mussina and Persons, 1963; and Bourne and Bunderson, 1963, for example) generally reported no significant difference between the effectiveness of immediate and delayed feedback. Two exceptions were Greenspoon and Foreman (1956) and Bourne and Bunderson (1963), which showed feedback as being more effective when it was delayed. As pointed out by Bilodeau (1966), it was normal to find confounded

variables in these early studies owing to the fact that little effort was made to control the three time intervals between initial response - KR - and post-KR response in multi-trial experiments. In addition, little effort was made to control what took place within the subject during these intervals of time.

More recent studies (Buchwald, 1967, 1969; Sassenrath and Yonge, 1969; More, 1969; Kulhavy and Anderson, 1970, 1972; Buchwald and Meager, 1974; Sassenrath, 1975; Surber and Anderson, 1975; and Peeck and Tillema, 1979) have more frequently shown delayed feedback as being superior to immediate feedback. However, the literature is still reporting similarly designed, recent studies (Beck and Lindsey, 1979; and O'Neil Rasor and Bartz, 1979, for example) which have found no difference between these two types of feedback. In addition, studies like White (1968) and Sassenrath, Yonge and Shrable((1968) reported that feedback was more effective when it was given immediately than when it was delayed.

Sevaral attempts were made to explain the trend that delayed feedback was at least as effective as immediate feedback in influencing performance (see Bilodeau, 1966; Sassenrath and Yonge, 1968, 1969; and Sturgis, 1969). At present, the most popular of these theories is an interference hypothesis proposed by Bourne (1966), to be developed later by Anderson and his associates cited above. Bourne (1966) concluded that as long as the time interval between the response and the feedback is not filled with interfering activities, learning is not affected by

delay of feedback. Supporting this view, Buchwald and Meager's (1974) delayed feedback treatment group actually made higher performance gains than the immediate feedback group, but only when subjects in the former group remembered their original responses.

Kulhavy and Anderson (1972) and Surber and Anderson (1975) reported convincing evidence that feedback does not act as a reinforcer. Their theory, referred to as the interference - perseveration hypothesis, postulates that during the time interval between the initial trial (t_1) and the delayed feedback, t_1 responses are not forgotten as readily. They perseverate; and proactive interference or response competition occurs when learning the correct response from feedback. Consequently, proactive interference should be strongest on t_2 test items which are most similar to test items on t_1. Their research (Surber and Anderson, 1975) showed that learning the correct response on a multiple choice test was better when negative feedback was delayed rather than given immediately. This was not so for positive feedback.

This finding seems consistent with their theory that the selection of any particular response to a test item strengthens the tendency to make that response. Where the initial response is incorrect, the strengthened tendency to repeat it interferes with the feedback information about the correct response. Since all responses weaken over time, delayed negative feedback is less affected by proactive interference than immediate negative feedback. Positive feedback, on the other

hand, is not interfered with whether given immediately or delayed. Therefore, the delay is less of an issue. In sum, then, research in this area suggest that an explanation based on mere S-R connections is too simple to illustrate how feedback affects behaviour in human beings.

Studies in linear programming. Linear programming is a form of teaching based on Skinner's (1953, 1954, 1958) view of operant conditioning. The programmes are designed to break down each unit material, 'A', into its various components, a_1, a_2, a_3, - - -, and to present them to the learner in a logical sequence. Tests are introduced at several well spaced points in the programme; and the items are designed to produce a very low error rate such that the correct, rather than the incorrect, responses will be reinforced or confirmed. This last point is central to Skinner's assertion that it is the immediate confirmation of the correct response which produces learning.

Mayer (1960) reported that immediate self-scoring in a linear spelling programme produced better results than not having the correct answer and leaving the scoring to the teacher. However, the error rate of his subjects' responses was quite high (14 per cent). Thus it was argued that the study was more of a Pressey-type, with KR serving to provide the subject with additional information rather than to reinforce or confirm the correct response. Almost every subsequent study (Moore and Smith, 1961; Feldhusen and Birt, 1962; Becker, 1964; Oppenhein, 1964; Dannemberg, 1965; Jacobs and Kulhani, 1966; and Barringer and Gholston, 1979) failed to obtain significant differences between reinforced and

unreinforced practice in linear programmes having a low error rate. Several of these studies included programmes in mathematics.

Guidance and prompting. If an information hypothesis of learning is to be valid, the test must include models which are independent of the time of arrival of the relevant information. Thus far information has been connected with KR and, consequently, has been restricted to post-response interpretations. One is curious to know whether this information would be at least as effective if made to precede the response rather than given as a consequence of action.

In multi-trial learning experiments, the result given after any one response also comes before the next trial. In these situations, then, KR comes between two trials. There is no reason why one should not envisage a model which stresses the practice effect of KR on the next trial.

Techniques used to examine the proactive effect of information on learning involve: i) the forced response where the learner is literally put through the motions; ii) visual guidance where supplementary information not contingent upon previous responses is given; and iii) verbal guidance which involves actually telling the subject what to do; in other words, directly prompting the correct response.

The typical experiment here involves giving the subject a few guided trials, say five to ten, and then requiring that he transfers

what he has learnt to unguided situations. The effectiveness of these exercises may be tested by the savings method, which simply estimates the amount of free practice still required for complete learning.

Gordon (1968), using a rotary pursuit task, compared three conditions: simple tracking, mechanically restricted tracking in which subjects chased the target by means of a stylus in a groove which permitted only lead and lag errors, and an augmented feedback condition in which a yellow light illuminated the field when S was on target. Mechanical guidance produced fastest initial learning but poorest transfer to unguided tasks. However, earlier studies (von Wright, 1957; Holding, 1959, for example) showed that the problem of poor transfer does not occur in guidance and prompting studies involving higher order learning situations. In their maze learning and paired associate learning situations, prompting and guidance treatments were as good as, or even superior to treatments involving KR type information given after the response.

Annett (1969) concluded from a review of verbal learning studies that prompting and guidance were normally equal to, and sometimes superior to, KR and that there was no reason to believe that the same was not true for studies involving perceptual identification and discrimination.

In sum, what little research done in this area supports the view that an information based interpretation of learning should be given credence over an explanation based on pure reinforcement.

Incentives and KR

The literature has associated at least two types of motivational effects with KR. These are a reinforcement effect and an incentive effect. One line of distinction between these two effects might be in terms of the permanence of the resulting change in behaviour. Reinforcement in KR has been associated with the production of more or less permanent behaviour changes, and with memory or retention of material. Incentive, on the other hand, has been attributed with little permanence, and has to do with speed, effort, and accuracy in some cases. Thus, whereas reinforcement has been seen by many theorists as being drive reducing, incentive has been seen to increase drive - meaning that it makes the subject expend more effort or work harder.

There are few studies on incentives involving tasks related to school disciplines such as arithmetic. Tasks used in these studies are usually very simple; and the dependent variables have more to do with the maintenance of performance levels than with the acquisition of skills.

Mace (1935) had subjects perform routine calculations. Subjects were given specific instructions about the targets they should attempt to achieve. All subjects received information on the number of computations they performed each session but were given different instructions concerning their performance targets. Students in one group were told simply to do their best. Another group was given a target of seventy computations per ten-minute period, and a third group was given the

target of exceeding the previous day's performance. The main results showed that subjects told to do their best improved faster and reached a higher level in twenty days than those given an absolute performance standard, and these did better than subjects told to better their previous performance.

Chapanis (1964), however, had subjects prepare computer tapes of random numbers and failed to find an incentive effect. His study involved four treatment groups. Group 1 was given no indication of the amount of work required or the amount completed; Group 2 had a counter recording their work but attention was not specially drawn to the counter, and it was never reset; Group 3 had the counter reset to zero at the beginning of each one-hour session; and Group 4 were asked to record the counter reading every 15 minutes, but were told that this information was for accounting purposes only. The relevant performance measure was in terms of work output, though accuracy was a secondary measure. Performance in all four groups improved over the 24 hour period, and there was no difference between the groups receiving different amounts of information. This was the only study reported showing no indication of an incentive effect. However, contrary to Chapanis' claims, the study seems to have no definite incentive treatment either.

Payne and Hauty (1955), in a series of exhaustive studies, gave various types of knowledge of results with the intention of differentiating between the informational and motivational effects of KR on the same task. Whereas both types of KR were associated with increased

level of performance, subjects in the incentive treatments showed a decline in performance level shortly after the first hour but informative KR did not result in work decrements. They concluded, therefore, that whereas informative KR continued to increase level of performance, incentive KR postponed decrement rather than abolished it.

Further research (Locke, 1966a, 1966b; Locke and Bryan, 1966a, 1966b) which included a repeat of Mace's (1935) experiment, showed productivity in situations involving incentive KR was more or less a linear function of the subject's intention or performance targets. Locke and Bryan's (1966b) three treatment groups were asked to perform complex computations for six successive ten-minute periods. Two groups received knowledge of the number of correct computations done in each period and scored these on their work sheets, but one of these groups was instructed on each trial to get a score of 15 more than their previous best. A third group was given minimal KR by listening to the correct answers being read by the experimenter but were not required to check their score or to better their performance. All groups were questioned after the experiment about the goals they set themselves. Though the three conditions did not differentially affect output, subjective goals did. The output was highest for those attempting to do their best and those simply aiming to improve their performance, with poorly motivated students producing a little more than half the number of calculations done by the highly motivated group. The results, therefore, suggest a difference between having a standard and simply having KR.

However, in the classroom learning situation, KR is usually the means by which the standard is specified, and by which performance is related to the standard. The question as to whether any effects associated with incentives are basically different or separate from effects associated with the informational aspects of KR has, therefore, not been carefully investigated. The conclusion reached by Annett's (1969) review of the relevant literature that incentive effects of KR are not basically different from informational effects is, consequently, left unchallenged. In sum, the research does not without question support the existence of an incentive effect of KR which is distinct from effects attributable to informational inputs.

Academic Self-Concept

Definition

Most self theorists agree that the self is a complex concept comprised of various feelings, attitudes and percepts one has for one's self. As noted by Wylie (1968), the hypothesized processes include perceptual discrimination, memorizing, verbal learning, concept formation, and symbolic problem solving. One's behaviours as remembered and as initially experienced are included in this conception (Rogers, 1969). The various percepts, concepts and experiences are assumed to be organized in a manner which possess some form of unity and a hierarchy determined by differing response strengths (Purkey, 1970; McClelland, 1951).

Some theorists (Purkey, 1970; Mead, 1934; for example) postulate semi-autonomous subsystems or subdivisions of the self. It is under this premise that Brookover and his associates (Brookover, Patterson, and Thomas, 1964; Brookover, Erickson and Joiner, 1967; and Brookover, Le Pere, Hamachek and Erickson, 1965) used the term 'academic self-concept', meaning self concept of academic ability. The construct includes the individual's self-evaluations, memories of his performances and experiences, and his perceptions of his competencies as compared with those of his age-peers.

Effects of Self Concept on Behaviour

Self theorists hold the view that behaviour is, to a recognizable extent, determined by the self concept. Snygg and Combs (1959), for example, state that all behaviour is determined by one's phenomenal field, the center of which is the self. Rogers' (1951) view is more widely referred to. He theorizes that it is necessary for two systems to function in any behaviour. The first of these is the person's self concept and perception of the conjoint world; the other is the learned and unlearned organismic processes and automatic behaviour. He goes further to state that the self regulates behaviour, being constantly used as a frame of reference (p.191).

In their effort to clarify the causal role of the self concept, some behavioural scientists (Brookover, Thomas and Patterson, 1964; Brookover, Erickson and Joiner, 1967) propose that self-concept enhancement can be a viable means of improving academic achievement. They believe that once the self is firmly established, the individual strives to behave in a manner that is consistent with it. Consequently, definite efforts to enhance the self concept will cause a comparable change in the subject's level of performance. However, models proposed by other social learning theorists (Secord and Backman, 1964, 1976) suggest that substantial and permanent changes in behaviour are not easily effected since there are many congruency processes in the self system promoting stability rather than change over time.

Research on Self Concept

Empirical evidence supporting claims of a causal role of self concept has lagged behind its incorporation into theory. Very many studies (Backman and Secord, 1968; Trowbridge, 1972; Rosenberg and Simmons, 1973; and Black, 1974, for example) have reported significant positive relationships between self concept and behaviour. These works, however, focussed on the developmental aspects rather than the causal role of self concept. Even the widely referred to study of Coopersmith (1967) had to do with the formation of the self.

Much of the correlational research, which includes longitudinal and cross sectional data (Brookover, Thomas and Patterson, 1964; Brookover, Erickson and Joiner, 1967; Epps, 1969; Reckless , Dinitz and Kay, 1957; Reckless and Dinitz, 1972; for example) has included measures of academic achievement, and has reported strong positive relationships between the achievement measures and self concept. For example, Brookover et al (1964) reported correlations of .42 (among high school boys) and .39 (among high school girls) between grade point average and self concept of ability, after controlling for measured intelligence. They used considerably specific definitions of self concept, viz self concept of ability in mathematics, english, science, and social studies. Their measures included the subject's perceptions of his abilities compared to his friends' classmates' and age-peers', as well as his parents', teacher's and peers' expectations of him. These definitions were in close consistency with the theory as proposed by Mead (1934).

Findings of this nature, however, do not provide evidence that self concept influences achievement.

Caslyn and Kenny (1977) reexamined the data reported by Brookover and his associates (Brookover, Le Pere, Hamachek and Erickson, 1965; Brookover, Erickson and Joiner, 1967), using cross-lagged panel correlational analysis. The data were derived from a longitudinal study of 556 adolescents. They noted that the correlations of achievement at time-one with self concept at time-two were higher than the correlations of self concept at time-one with achievement at time-two. This finding seems to support the view that self concept is an outcome of achievement rather than an intervening or causal variable. However, this does not rule out the possibility that the construct operates either way at different times or stages of the individual's psychic development.

The effects of self concept on achievement have also been examined in laboratory studies. These research efforts, however, have not consistently shown self concept as having a positive effect on academic achievement (Kraut, 1979). Very few, if any, of these studies have satisfied the methodological requirements specified by Campbell (1963), and recent reviews of the relevant literature (Rogers, Smith and Coleman, 1978; and Kraut, 1979) have reported that the negative and weakest effects are most common in conceptually weak investigations, which usually assume that the self-concept/achievement relation is independent of other environmental and psychological factors.

In an effort to confirm this conclusion, Rogers et al (1978) and Strang, Smith and Rogers (1978) conducted experiments using a stringent social-comparison model as advocated by Festinger (1954). They reported highly significant effects on every variable of interest, in the direction expected, though in many previous instances these effects did not reach significance under looser models. Their studies involved elementary school children, and the academic subjects were reading and arithmetic. Boersma, Chapman and Battle (1979) reported findings which supported the position of Rogers and his associates, using learning disabled, educable mentally handicapped, and normal elementary school children. Though the results of these studies are quite impressive, it seems unlikely that the persistent failure of previous studies to find significant effects is attributable to faulty methodology alone. There have been too many different measures taken on too many types of programmes to believe that important changes were being masked by weak research measures and designs.

In spite of its problems, one has to mention the Follow Through Planned Variations Project, since it dealt specifically with primary grade school children. Several types of evaluation of Follow Through were of national scope. Thye include the Stanford Research Institute's observational study (Stalling and Kaskowitz, 1974; Stalling, 1975) and abt Associates analyses of the overall model effects (Stebbins, St. Pierre, Proper, Anderson and Cerva, 1977).

The Stalling and Kaskowitz (1974) research focussed on the degree of implementation of seven of the model evaluation programmes. These involved 342 classes. The models differed strongly in emphasis given to the development of positive self concepts, and ranged from a strictly behavioural approach - which assumed that self concept, if it is ever a factor of importance, is a consequence of academic success - to the open classroom, which stressed that positive self concept is a necessary prerequisite of learning.

Final results of standardized tests revealed differences among programme sponsors favouring the behavioural approach on both academic achievement and self concept (Stebbins et al, 1977). This massive research effort, then, does not support the assumption that self concept is a causal or intervening variable of increased level of performance. A meta evaluation of the Stebbins et al (1977) analyses (House, Glass, Mclean and Walker, 1978) raised questions about the validity of their conclusions. Among other things, it was pointed out that subjects in the majority of models were not randomly assigned to experimental and control groups at the start of the project. The unsupportive results, therefore, may be attributable to weakness in the experimental designs.

In sum, one cannot conclude from the available empirical evidence that self concept is a necessary prerequisite of learning. Much of the research evidence seems to contradict the causality hypothesis. Other well designed studies (Rogers et al, 1978; Boersma et al, 1979), however, suggest that self concept has a positive effect on achievement

in school disciplines (mathematics and reading, for example) for under achieving primary grade school children, at least in some situations. The unanswered question is whether the observed differences could have been effected by other processes in the system of social interactions.

One can state with greater confidence that there is a positive relationship between self concept of academic ability and academic achievement; and the correlations tend to be highest with mathematics (Brookover et al, 1964, 1967; Rogers et al, 1978; Strang et al, 1978). This is a sufficient condition for the construct to be used effectively as a covariate in the present study.

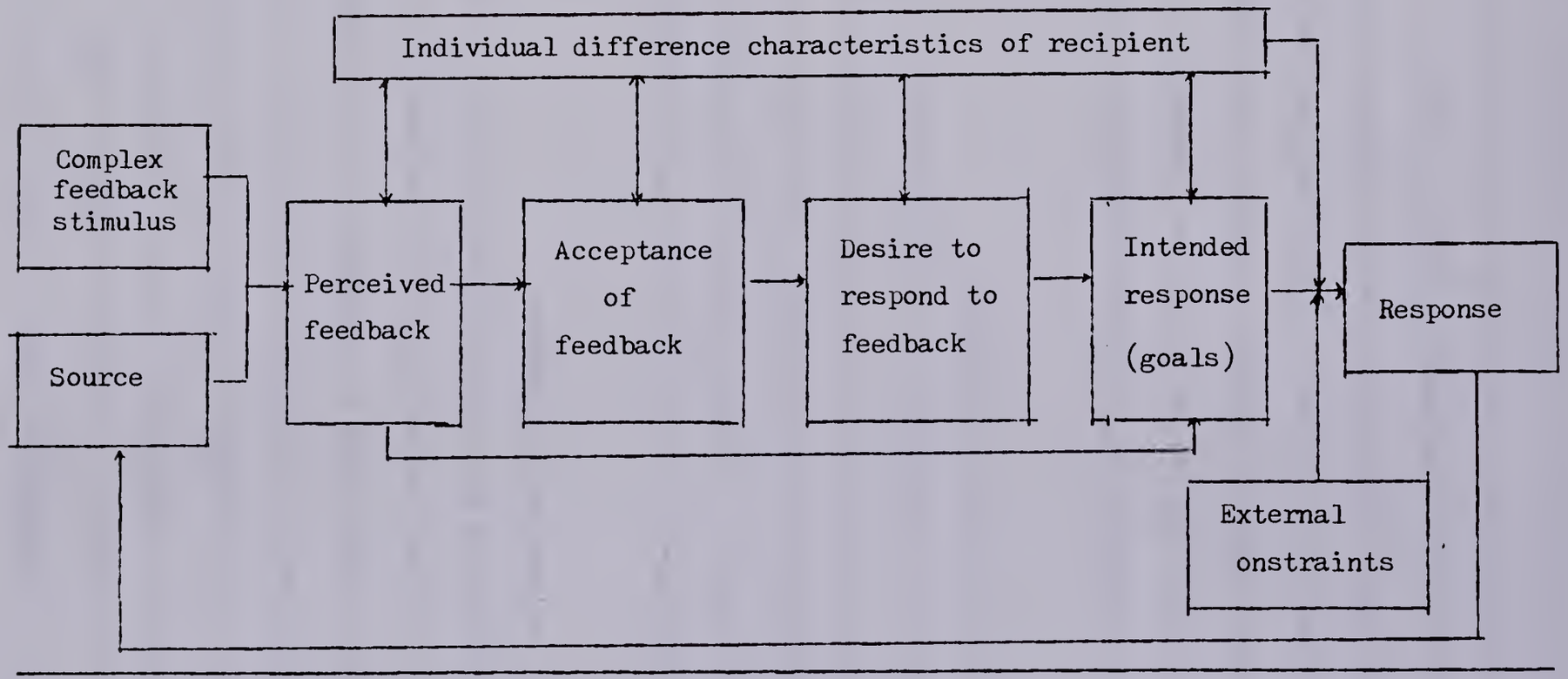
Performance Feedback as it Interacts With Self Concept, Feedback Sign and Feedback Preference

Figure 2.1 displays the elements of the feedback process. Feedback is shown to proceed from some source and is perceived, accepted, and related to targets or goals before influencing the recipient's response. At different points in the system, the individual's psychological characteristics - such as his frame of reference, which includes his self concept, locus of control, and cognitive set - influence the processing of the feedback stimulus, and consequently, partly determine the response.

Basically, feedback seems to be information received by the

Figure 2.I

Model of the Effects of Feedback on Recipient



(Source: Ilgen, Fisher and Taylor, 1979.

individual about his performance (Annett, 1969), indicating the correctness, accuracy, or adequacy of his response (Bourne, 1966). Its effectiveness must be judged subjectively since all perceptions are subjective. Besides depending on the increase in absolute value over information the subject already possesses, the recipient must be able to convert the feedback message into units that are meaningful to him (Bourne, 1966; Bilodeau, 1966). In other words, distinctive features of both the recipient and the stimulus must be considered if one is to gain insights into how feedback influences performance.

One characteristic of the recipient that seems important in this situation is self concept. Silverman (1964) reported that high self-esteem subjects remembered more information about a task in which they succeeded compared to one in which they failed. This finding vaguely suggests a tendency for individuals to be more receptive of feedback which is consistent with their self-regard.

Shrauger and Rosenberg (1970) further investigated this implication. They reported that high self esteem people, as compared to low self-esteem people, showed greater positive changes in their subsequent self-evaluations of a specific attribute following positive feedback on it, and that low self-esteem people showed greater negative changes in their self-ratings on an attribute following negative feedback on it. These findings are interesting since the two self esteem groups did not show significant initial differences in their self ratings with respect to

the particular trait on which feedback was given. Also, their findings showed negative tasks evaluations as producing substantially poorer subsequent performance only for low self concept students, and positive evaluations resulting in better performance only for high self-esteem individuals. Thus, in their experiment, feedback produced significant changes in performance only when it was consistent with the subject's level of self-ratings. The direction of change presented may be seen as being more impressive when one considers that it involved counteracting the regression-toward-the-mean phenomenon.

Silverman (1964) explained his findings as resulting from the individual's behaviour of giving less credence to situations where he has failed and paying greater attention to situations where he has been successful. However, the more recent study of Shrauger and Rosenberg (1970) suggested that, in addition to this selective valuing behaviour, the individual's performance changes following feedback may be mediated by the extent to which he has changed his self-evaluations in accordance with that feedback. A few other efforts (Kanfer, Karoly and Newman, 1974; and Korman, 1970; 1976) have supported these claims of an interaction between self concept and the effectiveness of feedback. Most of these research efforts involved adults in job oriented organizations. Consequently, one cannot be very conclusive about the presence of a similar interaction for students in an elementary school situation. The probability of such an interaction occurring, however, may be greater with a self concept measure that is more highly related to performance in school disciplines. Such a measure could be self concept of academic ability.

Feedback Sign

The extent to which the recipient accepts the feedback may be influenced significantly by its sign. Positive feedback seems to be more generally accepted than negative feedback (Jacobs, Jacobs, Feldman and Cavior, 1973 Halperin, Snyder, Shenkel and Houston, 1976). In the Halperin et al(1976) study, whereas positive feedback was generally accepted irrespective of its source, negative feedback was accepted only when it came from a high status source. The most popular explanation for this tendency has to do with the operation of defense mechanisms and the self concept. Thus positive feedback tends to be perceived and recalled more accurately, whereas negative feedback may be denied (Greller and Herold, 1977; Ilgen and Hamstra, 1972).

Preference for Type of Feedback

This section deals particularly with the interaction of feedback preference and feedback delay. No research has been done in this area with human subjects. The studies with animals are quite decisive, and consequently, merit some consideration here.

Davenport (1962) reported that larger rewards offset loss of interest caused by longer delays. His rats' preference took an interesting turn, however, as the experiment progressed. After 72 trials, the rats shifted to favour four pellets at a 30-second delay when this alternative was pitted against 2 pellets at a 1-second delay. And yet, early in

the experiment, they went for the delayed larger reward 75 per cent of the time. Logan (1968) confirmed these findings with another set of rats. The consistency of the findings reported in replications of this study give some prominence to the view that there is a separate urge to have the reward immediately irrespective of its size or capacity for need satisfaction.

The study of Ainslie (1974) revealed further insights into the impulse to select the immediate reward. His study involved giving pigeons a choice between pecking at a key that would deliver a small reward almost immediately, or pecking at another disc that would deliver a larger reward after a longer delay. His pigeons virtually always choose the small reward delivered immediately. However, when the pigeons were given the opportunity to shut off the smaller reward if they wished, they almost always eliminated the option of getting a smaller reward immediately and chose the larger reward after the longer delay. In other words, they seemed able to choose the larger reward after the longer delay only after getting rid of the possibility of being rewarded immediately. Ainslie explained the initial behaviour of the birds to select the immediate reward as impulsive or automatic, since some birds took the immediate offer and then waited for the delayed reward. The point of greatest relevance to the present study, however, is the suggestion that there are two distinct forces operative in these situations. They are the immediacy of the reward and the size or capacity of the reward to satisfy the individual's need. Th research also suggest that these forces may work against or counteract one another.

III

THE DESIGN

The present study was designed to examine the influences of two types of feedback on performance in arithmetic, as well as their interaction with self concept of ability in arithmetic. As stated in an earlier chapter, the two feedback treatments may be referred to as knowledge of results issued immediately on completion of each test form, and a similar feedback administered one day after subjects had completed each test. The former feedback treatment, then, was really immediate posttest knowledge of results (IKR), where the latter was termed 'delayed knowledge of results' (DKR).

Information feedback was issued on the student's performance with respect to three test forms described in detail in a subsequent chapter. The testing-feedback machinery was designed to carry out three main functions in this study. These may be stated as: i) to provide practice or drill on arithmetic skills pertinent to the student's curriculum; ii) to serve as a mode of instruction whereby students could gain relevant information about, or insights into, the right answer and their likely mistakes; and iii) to provide a measure of the student's ability to carry out the skills involved. It was thought that the student's performance would improve once their attempts at doing the tests were followed by performance feedback.

The study took on a repeated measures design with different subjects in each cell, so as to eliminate cross over effects from one treatment to another. The cells were defined by three treatments: IKR, DKR, and no-KR, and three occasions. The main aspect of the study, then may be seen as a 3(treatments) X 3(occasions) factorial experiment with subjects nested in treatments. Self concept of ability in arithmetic and other significant self concept concomitants of performance in arithmetic were used as covariates.

The Main Questions

The analyses undertaken were guided by three main questions:

1. With respect to the criterion, performance in arithmetic, are the population mean profiles over the three occasions similar for the three treatment groups? Put in another form, are the line segments defined by the test means for each treatment group parallel?

Here, the aim was to test the hypothesis of no group-by-time interaction. This interaction was expected to be significant, meaning that as the experiment progressed from Occasion 1 through Occasion 3, the treatment groups were expected to increase their level of performance at different rates.

The rationale supporting this proposition has to do with the general

belief that learning increases with time and added experience. As time goes by, people become exposed to new experiences and acquire added skills which eventually influence their performance. Though the time spanning the experiment was short and roughly constant across treatments, each group was exposed to different experiences and, consequently, should show different rates of learning.

2. The second question was designed to examine differences between the KR/treatment effects. It may be stated as follows:

> If the population mean profiles for the two KR treatment groups across the three occasions, are indeed parallel, are they also on the same level; that is, do they fall on one another?

An answer in the affirmative was expected, meaning that the delayed KR treatment was expected to be at least as effective as the immediate posttest KR treatment.

Several assumptions were made in conjunction with this proposition. Firstly, in accordance with the interference-perseveration hypothesis as proposed by Anderson and his associates (Kulhavy and Anderson, 1972; Surber and Anderson, 1975), it was assumed that in situations where students' initial responses were incorrect and feedback was given immediately, the correct and incorrect answers to the stimulus question would be more similar in strength than in situations where this

feedback was delayed. Consequently, there would be greater interference (accompanied by poorer performance) in the immediate feedback situation than in a situation where feedback was delayed. However, the fact that several studies (Beck and Lindsey, 1979; Newman, Williams and Hiller, 1974; Phye and Baller, 1970, for example) reported no difference between these two feedback treatments, and others (Sassenrath and Yonge, 1968; White, 1968, for example) found immediate feedback to be the better mode of KR, suggest that other factors may be operative in these situations determining the relative efficacy of these two types of feedback. One such factor may be the nature, in terms of similarity or dissimilarity, of interpolated activities (Bilodeau, 1966; Bourne, 1966). Others may stem from the influence of the individual's psychological idiosyncrasies such as academic self concept and mental energy which can influence the frequency of rehersals and the amount of energy brought to bear on the learning situation.

Also, a recent study (O'Neil, Rasor and Bartz, 1979) raised questions as to how long a delay was necessary to offset detrimental effects of learning caused by proactive inhibition. They examined differences between two types of immediate item-per-item feedback and immediate posttest feedback. The immediate posttest feedback was clearly better ($p < .05$) than the other two treatments; and the level of significance was similar to that reported in most studies which suggest that delayed feedback is superior to immediate feedback. The inconsistency of the findings reported in the related literature, therefore, does not support the proposition that either one of these two feedback treatments (IKR

and DKR) should be more effective than the other.

3. The third question was designed to examine the overall effects of KR on performance in arithmetic. It may be stated as follows:

> Are the population means for the treatment groups different from that of the control group? In other words, is the pooled mean of the groups treated with KR different from that of the group which did not receive feedback?

It was expected that KR would have a positive effect on grade four students' performance in arithmetic. It was thought, therefore, that the average performance gains of the groups treated with KR would be higher than the corresponding gains of the no-KR treatment group.

With respect to this assertion that KR should positively influence the students' performance in arithmetic, it was assumed that grade four students internalize goals or expectations of high performance their parents and teacher hold of them, and that they perceive the information in feedback as being useful to them in terms of specifying the means by which they could better their performance. Coopersmith (1967) and Newman et al (1974) serve as support to these assumptions.

Findings in the direction expected would support claims made by Locke (1967), O'Neil, Rasor and Bartz (1979), and Beck and Lindsey (1979) that feedback produced positive effects on performance. It

should be noted, however, that no effort was made in this study to eliminate the presence of goals or targets the students may have set themselves. Consequently, findings in the direction expected would not necessarily support claims that it is the informative aspect, rather than any associate targets, of feedback that influence performance.

The Interaction of Feedback and Self Concept

This aspect of the study was designed to examine the interaction between certain measures of self concept of academic ability and the effectiveness of the feedback treatments. Its purpose, therefore, was to discover which treatment was best for students with various levels of self concept. Also, as a subsidiary goal, this section examined the relationship between self concept of ability in arithmetic and performance in arithmetic as the dependent variable. The latter measure was used as a checking device rather than as a major finding of the study. Instead of using self concept of ability as a stratification or blocking variable, it was decided to use the construct as a concomitant in a series of covariance analyses for several reasons: Firstly, the criteria on which the sample was selected did not satisfy the specifications of a randomized blocks design. Secondly, it was thought that the covariation between self concept of ability and the criterion was high enough to produce a gain in the reduction of the error term via covariance analysis which was nearly as high as the corresponding gain derivable from the use of the construct as a blocking variable. And

thirdly, covariance analyses allow for greater power in assessing the significance of interactions through their incorporation of regression analyses.

It was expected that self concept of ability in arithmetic would show a positive regression slope with performance in arithmetic. That is, students low in self concept would show lower levels of performance than would students high in this construct.

Several correlational studies reported in the literature review indicated findings which lend some support to this expectation. All of these studies reported correlation coefficients falling within the range of .42 and .25. Regression coefficients computed in several instances fit a similar range. There is no supported reason to expect different regression estimates in the present study.

Another expectation has to do more directly with the interaction of self concept of academic ability and the effectiveness of the treatments.

A significant treatment-by-self-concept interaction was predicted. More precisely, it was thought that a higher regression estimate would be observed for the IKR than for the DKR treatment group.

Here, it was assumed that the problem of response interference

or competition would be more of an issue with the IKR than with the DKR treatment group. It was further thought that subjects high in self concept would approach the task with greater determination, exerting more effort to achieve higher grades. Consequently, they would be more likely to overcome hindrances caused by interference processes. High self concept students, then, may perform well under both IKR and DKR treatments, but the difference in performance may be more remarkable for low self concept students. The prediction, however, is left hanging without empirical evidence to stand on.

Preference for IKR or DKR

No study involving human subjects has been conducted to link preference for IKR or DKR with the effectiveness of these treatments on level of performance. However, the studies done with animals (Davenport, 1962; Logan, 1968; Ainslie, 1974; D'Amato and Cox, 1976, for example) suggest a strong preference for the immediate reward. These research efforts focussed on the preference for a small immediate reward against a delayed larger reward, and have been fairly consistent with the finding that the size of the reward could offset the negative effects of delay.

However, the studies mentioned above extended their investigations over a longer period of time and indicated an undercurrent impulse to select the immediate reward irrespective of its comparative size. Less intelligent subjects like rats would, after a while, select the

immediate reward and then wait for the delayed reward, whereas the more intelligent subjects like pigeons could select the delayed reward only after getting rid of the stimulus to have the reward immediately.

Other variables may be operative in parallel situations involving human subjects. Proactive inhibition or response competition may be one of these factors in retention and recall learning situations. If this is the case, students may prefer a conflict free situation to one which fosters response competition, and consequently , choose to have feedback on their performance delayed rather than immediate. On the other hand, where feedback is seen to have rewarding elements, the urge to have the reward as soon as possible may work against other desireable features, like reward size, and serve to make immediate posttest KR and one-day delayed KR equally attractive to students in this population. However, as this feedback is delayed for a longer period, one week for example, a definite preference might be shown for the immediate reward. The lack of empirical evidence in this area of research made it difficult to predict which of these forces would be stronger, and consequently, which of these types of feedback would be most attractive to students in this population. Any finding that immediate posttest feedback and one-day delayed feedback are equally desireable, however, would on the surface seem to be consistent with the interference-perseveration hypothesis.

IV

SAMPLE, INSTRUMENT, AND PROCEDURE

The Sample

Grade four students in four schools governed by the Edmonton Public School Board served as subjects of this study. The schools were not chosen on a purely random basis. Rather, as determined by the Edmonton Public Board, criteria for selection had to do with:

i) the pupil's previous involvement in similar projects,

ii) the school's availability to outside researchers, and

iii) the principal's willingness to become involved in a project of this nature.

Together, these schools had 276 students in eleven Grade 4 classes. Of these, 243 students in 9 classes served as subjects of the experiment. Seven of these students were absent from one or more parts of the experiment, and consequently, were dropped from the analyses. The final sample, therefore, was comprised of 236 Grade 4 students, of which 117 were females and 119 males. Their ages ranged from 7 years 11 months to 11 years 8 months, with a mean of 9 years 4 months. 230 of these students fell between the narrower age-range of 8 years 9 months to 10 years 7 months. Students in the other two classes were used in a pilot study described later in this chapter.

The Instruments

Performance in Arithmetic

Three forms of an arithmetic achievement test were used to measure performance in arithmetic. As stated earlier, the tests were designed to provide the students with practice on arithmetic content that were taught previously. Therefore, they may be described as tests of arithmetic operations.

Initially, the overall test domain was framed to include selections from Grade 2 to Grade 3 curriculum guide for teachers in the province of Alberta. These selections had to do with addition and subtraction involving two or three digit whole numbers, with or without regrouping. Any multiplication or division dealt exclusively with the basic facts. The domain was segmented into little subdomains, one to correspond with each potential test item. Each domain statement contained specifications of content as well as of other features such as format and item generation scheme (see Appendix A).

With the aid of random number generators, the computer was used to create items in sets of three, one for each form of the test. Numbers were generated three at a time from each content specification, 101 to 999, for example. For each item, this process was repeated until the features of the particular subdomain were satisfied. Item format was held constant for each trio.

The Instruments

Performance In Arithmetic

Three forms of an arithmetic achievement test were used to measure performance in arithmetic. As stated earlier, the tests were designed to provide the students with practice on arithmetic content that were taught previously. Therefore, they may be described as tests of arithmetic operations.

Initially, the overall test domain was framed to include selections from Grade 2 and Grade 3 curriculum guides for teachers in the province of Alberta. These selections had to do with addition and subtraction involving two or three digit whole numbers, with or without regrouping. Any multiplication or division dealt exclusively with the basic facts. The domain was segmented into little subdomains, one to correspond with each potential test item. Each domain statement contained specifications of content as well as of other features such as format and item generation scheme (see Appendix A).

With the aid of random number generators, the computer was used to create items in sets of three, one for each form of the test. Numbers were generated three at a time from each content specification, 101 to 999 for example. For each item, this process was repeated until the features of the particular subdomain were satisfied. Item format was held constant for each trio.

Preliminary forms of the achievement test, and the pilot study.

Using random number generators as noted above, three preliminary forms of the arithmetic achievement test were prepared for the first trial run. Each form contained 28 items, composed for the most part of numbers and mathematical symbols. Each class teacher involved in the project was asked to rate each item in terms of its difficulty and familiarity to her pupils.

One of the eleven classes was chosen at random for the first trial run. The exercise took the form specified by Angoff (1968) where two test forms may be considered equivalent if they show a similar practice effect on a third replica of this test. Forms 1 and 3 were chosen at random to be administered during the first session of this pilot study. Alternating one with other, both test forms were stacked in one pile and then distributed from the top downwards such that the class was divided into two subsections.

The second session was taken the same morning, one hour after Session 1 was completed. Subjects who had responded previously to Form 1 took the third form, whereas those who had written Form 3 in the first session responded to Form 1 in Session 2. A third session was to follow. However, this part of the study was dropped since the results obtained from Sessions 1 and 2 made a third session redundant.

The two test forms correlated positively with each other, $r_{xy} = .47$, though a much higher covariance/variance ratio was desired. The items

were considered too easy, difficulty levels ranging from .39 to 1.00, inclusive. The means were 21.9 and 22.8 for Forms 1 and 3, respectively.

Re-defining the domain. Attempts at obtaining items closer to the .5 difficulty level and a higher covariation of test scores had to do with a re-definition of the total domain, as well as putting tighter restrictions on the subdomains of each test item. This strategy was preferred over the modification of items already produced, since: i) the modification of items randomly produced would affect their ability to refer to the domain, and consequently, their equivalence from a theoretical point of view; and ii) a re-definition of the domain better allowed for the production of the type of items the researcher wished to construct.

All selections from the Grade 2 curriculum guide were considered too easy for students in this population, and consequently, were excluded from the domain. All addition sums, therefore, had at least three addends, and those limited to numbers of fewer than three digits were restricted to horizontal arrangements. Other re-definitions took the form of removing one, zero, or multiples of ten from the field of some subdomains. In other instances, these numbers were dealt with separately.

The second trial run. Three forms of the achievement test, each of 28 four-option multiple choice items were constructed in the same way as in the first trial run. Students from one of the eleven classes were required to respond to the items without assistance other than that given in the directions printed on the second page of each test booklet.

As in the first study, care was taken to ensure that each student gained a good understanding of the directions and could record his answers as required on the answer sheet. An analysis of the responses indicated item difficulties ranging from .444 to .940 with 23 of the 28 indices distributed evenly between .667 and .889.

A further attempt was made to increase the difficulty levels of the easier items. This involved examining the types of wrong responses in order to identify popularly selected incorrect responses. Easily identifiable incorrect options were replaced by ones that would seem to be more likely selected by students in this population. In addition, two of the easiest items which had roughly zero discrimination indices were excluded from each of the final test forms.

The final forms of the arithmetic achievement test.

The final forms of the arithmetic achievement test were each composed of 26 four-option multiple choice items. The content, format, and other features of these items were as specified earlier. Item analyses were done on the first trial of all students in the total sample. The ranges of item difficulties were: i) .333 to .885, ii) .341 to .854, and iii) .312 to .896, for Form 1, Form 2 and Form 3, respectively. For each form, 22 of the 26 difficulty levels were distributed roughly evenly between .55 and .85. Similarly, the majority of the item discrimination indices were distributed without any bias between .33 and .70. None of the items was negatively discriminating, the lowest index being .2. For each form, all except two of the

item scores showed significant positive correlations with the total score, as indicated by the corrected point-biserial correlation coeffi. cients. For each test form, these two items showed positive correlations which were not significant from zero. The means, standard deviations, and standard errors of measurement are displayed in Table 4.1.

Table 4.1

Statistics of the Final Test Forms

Test form	n	Mean	S.D.	S.E.	Cronbach's alpha
Form 1	78	17.81	5.01	2.0757	.8284
Form 2	82	16.33	5.19	2.1952	.8212
Form 3	77	18.05	4.69	2.1003	.7992

Note. Maximum score = 26

Internal consistency and equal forms reliability. Table 4.1 displays the Cronbach's alpha coefficients for the three test forms. These indices were not substantially different and were acceptably high.

Equal forms reliability indices were computed from the intercorrelations of the three test forms. The observed correlations computed from data had from the control group (N = 72) were as follows:

$$r_{12} = .7268, \qquad r_{13} = .6705, \qquad \text{and } r_{23} = .8085.$$

Reliability indices computed by taking the square root of these coefficients were:

$$r_{xt_{12}} = .8525, \qquad r_{xt_{13}} = .8188, \qquad \text{and } r_{xt_{23}} = .8992.$$

The reliability of the three test forms taken as one (r_{ttt}) was computed with the aid of the formula:

$$r_{ttt} = 1 - S_e^2/S_t^2 \; ; \quad \text{where}$$

S_e^2 was defined as the weighted mean of the three squared standard errors; and S_t^2 as the weighted mean of the three variances.
This index was appreciably high, $r_{tt} = .8172$.

Test of equivalence. The t-test of difference between two dependent correlations was used to assess the significance of the difference between the observed correlations reported in the previous section. The observed t-values were: 1.124, 1.568, and 2.692, for $r_{12} - r_{13}$, $r_{12} - r_{23}$, and $r_{23} - r_{13}$, respectively. Two of these values were smaller than the critical t(70) at the .05 level (which was 1.997) suggesting that the corresponding differences were insignificant.

In order to consider together scores taken on the three test forms, further computations were carried out with the aid of a 1-way analysis of variance program. The Occasion 1 scores for the entire sample served as the dependent variable, whereas the three test forms were used as three levels of the independent variable.

A summary of the analyses displayed in Table 4.2 showed the three test forms as being not significantly different, the observed F-value being smaller than the critical $F(2, 234) = 3.04$ at the .05 level. Scheffe multiple comparisons between the means of the three test forms also showed all contrasts as being insignificant, $p > .10$.
The variance of the scores was homogeneous across the three test forms, $\chi^2_{obs.} = .909$; $p = .635$.

Table 4.2

1-Way ANOVA of Occasion 1 Scores on the Three Test Forms

Source	SS	df	F	Prob.
Groups	138.418	2	2.77	.065
Error	5840.859	234		

Scheffe Multiple Comparisons

	Mean difference			
Form 1 - Form 2	1.50	2, 234	1.81	.166
Form 1 - Form 3	- .19	2, 234	.03	.972
Form 2 - Form 3	-1.70	2, 234	2.29	.104

Note. All contrasts are insignificant at the .10 level.

See Table 4.1 for a display of the means and the n's.

At this point, it should be recalled that the test items were drawn at random from a well defined domain in a manner which allowed the test forms to satisfy the theoretical requirements for equivalence. The purpose of the following computations was merely to describe the statistical characteristics of the instrument.

Wilks (1946) method of examining the equivalence of three or more forms of a psychological instrument was used to assess the similarity of the three measures. The method included three stages: i) the equality of means, equality of variances, and equality of covariances test (L_{mvc}); ii) the equality of variances and equality of covariances test (L_{vc}); and iii) the equality of covariances test (L_m) -- parts two and three being executed only on the condition that the previous test(s) failed.

The formulas for the three stages of this examination, applicable to three test forms, are recorded in Gulliksen (1950; p. 181). The data taken on all three occasions from the control group was used for these computations. Table 4.3 displays the observed values and shows how they compare with the critical values as suggested by Wilks (1946; pp. 263-266). The observed L_m = .96493 which was greater than the critical values of .9513 and .9261 at the .05 and the .01 levels, respectively. Consequently, the observed $-\underline{n}(k - 1)\ln L_m$ of 5.1412 was smaller than the corresponding chi square critical values of 5.99 and 9.21, respectively. The three test forms, therefore, may be considered equivalent, allowing for adjustments on the variances and/or the means.

Table 4.3

Tests of Equality of Means, Variances, and Covariances

Source	Observed values	Critical values .05	Critical values .01	df
L_{mvc}	.77845	.8135	.7591	6
$-\underline{n}\ln L_{mvc}$	18.0322	12.5916*	16.8119	6
L_{vc}	.83607	.8549	.8029	4
$-\underline{n}\ln L_{vc}$	12.8910	9.4877	13.2767	4
L_{m}	.96493	.9513	.9261	2
$-\underline{n}(k-1)\ln L_{m}$	5.1412	5.99147**	9.21034	2

Note. df for $L_{mvc} = (k/2)(k+3) - 3$;

df for $L_{vc} = (k/2)(k+1) - 2$;

df for $L_{m} = k - 1$;

$k = 3$; $\underline{n} = 72$

* $-\underline{n}\ln L_{mvc}$ and $-\underline{n}\ln L_{vc}$ distribute themselves as chi squares for $\underline{n}$'s larger than 63 where $\underline{k} = 3$.

** $-\underline{n}(k-1)\ln L_{m}$ distributes itself as a chi square for $\underline{n}$'s larger than 61 where $\underline{k} = 3$.

Since the test of L_m (referred to as H_m) is comparable to an anlaysis of variance test, a further check on the test variances was made by running a one-way analysis of variance on this data (from the control group). The majority of the differences between the scores (as considered across the three test forms) was attributable to differences in variances, $p < .001$. However, as pointed out earlier, when similar analyses were done using the scores taken on Occasion 1 only, and all 9 classes rather than just the 3 classes in the control group, the variances observed were very similar, $p > .05$.

The Scorer

An electronic '3M Model 550 Test Scoring Computer' was used to score the students' responses to the achievement test. It necessitated that students record their answers on specially prepared answer sheets which could be easily fed through this machine. Each answer sheet used in this study could hold the responses for up to 100 items, 50 on each side of the sheet (see Appendix B).

The scorer allowed for prompt marking of the test responses. Once it was programmed with a key for the particular test form, each student could more to the scorer immediately on completion of the test, feed his answer sheet through, and obtain his results. Feedback provided included a small red dash along side each incorrect response, as well as a score indicating the number of items done correctly. A rescore feature was available, also, allowing students to obtain a second score after using the feedback provided in the first instance. This involved shading a rescore option at the foot of the answer sheet.

Self Concept of Academic Ability

The Student's Perception of Ability Scale (SPAS) (Boersma and Chapman, 1977; Boersma, Chapman and Maguire, 1979) was used to measure self concept of academic ability. The scale was constructed to measure the Grade 3 to Grade 6 child's perceptions and interactions related to his performance on academic and school related tasks, as distinct from those of other children his age. As such, then, the scale was specially designed to measure self concept of academic rather than general ability. Correlations ranging from .029 to .078 between SPAS and the 'Piers-Harris Children's Self Concept Scale' (Piers, 1969) clearly suggest independence of the two self concept measures.

The full instrument is contained in 70 short statements. They are all simple enough for the Grade 3 child to read and understand with very little, if any, assistance; and each requires only a 'YES' or 'NO' response. The instrument includes six subscales. Three of these are 'Perception of Arithmetic Ability' (12 items), 'Perception of Reading Ability' (12 items), and 'Perception of General Ability' (12 items). Correlations among the subscales and between the subscales and the full scale suggest that the subscales are relatively independent of each other, though together they measure a common construct - Cronbach's alpha for the full scale being .915. For the subscales, the alphas ranged from .686 to .855 with a median of .803. Test-retest reliability coefficients -- .834 for the full scale, and ranging from .714 to .824, with a median of .765, for the subscales -- are also high.

Procedure

Assigning Classes to Treatments

As stated earleir in this chapter, the experimental sample was contained in 9 classes. Three of these classes were in one school, whereas the other six classes were distributed evenly among three other schools.

Each of the three classes in the first school was randomly assigned to a separate treatment. The assignment of the other six classes to the treatments was also random, once the possibility that two classes in the same school fall in the same treatment group was exempted from the realm of favourable options. Each treatment group, therefore, was composed of three classes. The immediate KR treatment involved 87 subjects, whereas 84 and 72 subjects were in the delayed KR treatment and the control groups, respectively.

Administration of the Achievement Test

As specified in the previous chapter, each group was subjected to only one of the treatments so as to eliminate cross over effects from one treatment to another. In each school rather than completing the experiment with one class before taking another, all treatment groups started the exercise the same day so as to minimize differences in previous learning. Whenever it was feasible, the tests were written in the

scheduled mathematics periods so as to minimize effects of additional teaching between Occasion 1 and Occasion 2.

The order in which the three achievement test forms were presented was random for each treatment group. Consequently, one class in each treatment took the tests in the order: Form 1, Form 2, Form 3; another class wrote Form 2 first, then Form 3, and finally Form 1; whereas the third class wrote Form 3 first, then Form 2, and finally Form 1. This strategy was developed so as to let any error of measurement attributable to differences between the test forms or day of administration enter into the residual term of the model.

The IKR treatment group. The immediate posttest KR treatment group was required to respond to the three test forms on three different occasions. On Occasion 1, all students in this group were given a test booklet and an answer sheet. They were asked to write their names on the answer sheet, and then to follow the directions on the second page of the test while they were read by the investigator. In addition to specifying the method of responding to and purpose of the test, the directions required that each student do two sample items and record answers on a duplicate answer sheet. The students were then asked to begin the test. Care was taken to see that each student recorded his answer to Item 1 on the first row of the answer sheet, and his answer to Item 2 on the second row, and so on. Each student was given enough time to respond to all items of each test form.

Each student in the IKR treatment group was instructed to feed his answer sheet, upon completing the test, through the electronic scorer which was already programmed to mark this test form. Feedback given included a score which was the total number of correct responses, and a red dash beside each incorrect response. The student was also instructed to try for a second time at each item done incorrectly. This was to ensure that he received or payed attention to the feedback and to increase his chances of finding the correct answer in this period. Students who did not arrive at the correct answer on a third attempt were assured that the unselected option was the correct answer. In all instances, the first score was the one chosen for the analyses.

Occasion 2 for students in this treatment group was on the following day, that is, one day after receiving feedback on Occasion 1. On this occasion, these students were required to write another form of the achievement test and to obtain feedback in the same manner as on the previous occasion. Occasion 3 was one week (7 days) after Occasion 2. On this occasion, the students were required to write yet another form of the test and to score their responses as on the previous occasions.

The DKR treatment group. Students in the one-day delayed KR treatment group were required to write the test forms on three different occasions, as did students in the IKR treatment group. The directions, conditions under which the tests were written, and other features of the situation were similar to those of the IKR treatment group, except for the time of receiving feedback. Whereas students in the IKR group were

treated with feedback immediately after they had completed each form of the test, students in the DKR treatment did not receive their feedback until one day after. For this group, the answer sheets were collected as soon as the students had completed the test and were redistributed for marking the following day. Each student was required to check his own responses by feeding his answer sheet through the scorer. He was given a second chance to try at items done incorrectly on first attempt so as to ensure that he paid attention to the feedback. Scores obtained on second and third attempts were not used for the analyses. Here, as with the IKR treatment group, students who did not get the items correct by the third attempt were assured that the unselected option was the correct response.

Occasion 2 for students in the DKR treatment was one day after they had obtained feedback on Occasion 1. Here again, feedback was had one day after completing the test. A third attempt at the test (Occasion 3) was given one week (7 days) after students had received feedback on their second attempt at the test.

The control group. Students in this group wrote the three test forms on three different occasions, as students in the other two groups. Occasion 2 was given one day after Occasion 1, while Occasion 3 was had one week after Occasion 2. The features of the testing situation were similar to those of the other two groups. This group of students, however, did not receive feedback between trials. Rather, information was given about their performance on each of the three test forms on

Occasion 3 when they were allowed to score their results on the three test forms.

Students in all three treatment groups were administered the 'Student's Perception of Ability Scale' (Boersma and Chapman, 1977; Boersma, Chapman and Maguire, 1979) during the week between Occasion 2 and Occasion 3. Some class teachers offered to administer this instrument. Consequently, all class teachers were asked to administer the scale to their students.

In each treatment group, each student was required, also, to respond to one question investigating his feedback preference. The question was phrased as follows:

'If we decide to give you more practice exercises, when would you prefer to know your score:

i) IMMEDIATELY AFTER you have completed each test,

or ii) ONE DAY AFTER you have completed each test,

or iii) ONE WEEK AFTER you have completed each test?'

The question was put to the students after they were treated with KR such that their responses served as a post hoc measure of the experiment.

V

THE RESULTS

Overview

The results of the study are presented in three main sections. The first section deals with the effects of feedback on performance in arithmetic, and presents the analyses on the criterion scores without considering the influence of the self concept covariates. The second section investigates the influence of the concomitants, and consequently, takes up as its main concern the investigation between self concept of ability in arithmetic and the effectiveness of the different types of feedback. The final section considers the post hoc measures of the student's preference for immediate or delayed feedback. The analyses carried out in each section were guided by the research questions stated in a previous chapter.

The Effect of Feedback on Performance in Arithmetic

The data collected from the three treatment groups on all three occasions were analysed according to a two factor analysis of variance programme designed to handle repeated measures on one factor. The independent variables were the treatments and the occasions, and the scores obtained as measures of the students' performance in arithmetic on all three occasions were used as the dependent variable.

Table 5.1

Two Factor Analysis of Variance Relating Treatment and Occasion to Achievement

Source of variation	SS	df	MS	F	P	Sig.
Between subjects	14176.313	235				
Main effects	491.480	2	245.740	4.191	.016	+
Subjects within Grps.	13662.250	233	58.636			
Within subjects	4246.688	472				
Main effects	448.012	2	224.006	28.574	.000	+++
Interaction	121.890	4	30.472	3.887	.004	++
B X subj. within grps.	3653.250	466	7.840			

Note. + = significant at the 0.05 level

++ = significant at the 0.01 level

+++ = significant at the 0.001 level

The Treatment-by-Time Interaction

The question of interest here had to do with the significance of the interaction between the treatments and the occasions. This interaction was expected to be significant.

A summary of the analyses, Table 5.1, indicates a significant interaction, suggesting a differential effect of the treatments over time. The cell means are shown in Table 5.2 and are plotted in Figure 5.1. The observed F-value = 3.887, which was greater than the critical $_{.95}F(4, 466)$ of 2.23, and was very close to the conservative $_{.95}F(1, 77)$ of 3.96. The results, therefore, support the expectation of a significant treatment-by-time interaction effect.

Table 5.2

The Cell Means

Treatments	Occasions			Row means	n
	1	2	3		
IKR	15.951	18.222	18.568	17.5803	81
DKR	18.193	20.036	20.542	19.5903	83
no-KR	17.931	18.486	18.361	18.2593	72
Column means	17.3583	18.9157	19.1570	18.4767	

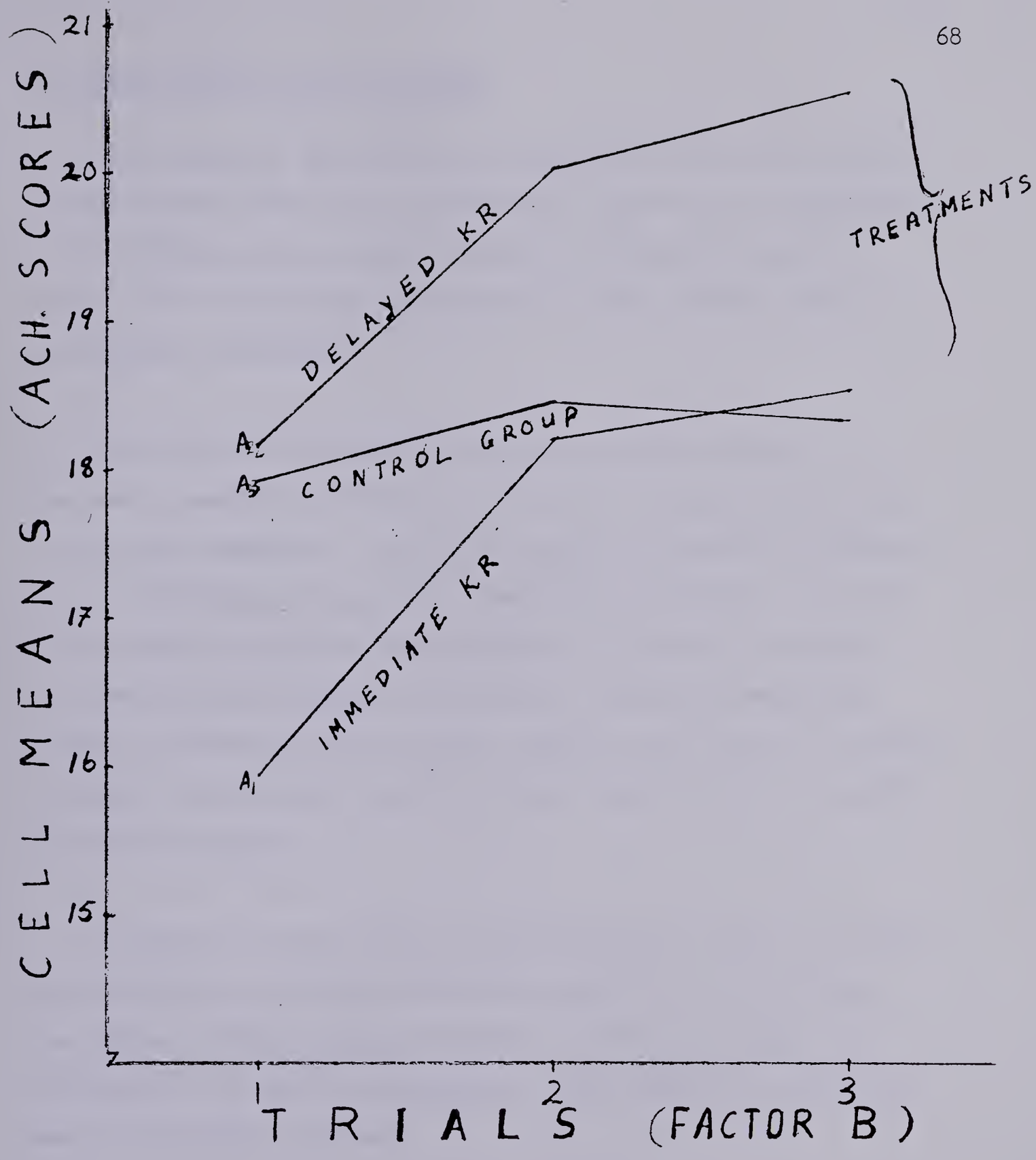


Figure 5.1
INCREASES IN PERFORMANCE OVER THREE TRIALS FOR THREE TREATMENT GROUPS

The Simple Effects of the Treatments

The analyses of this section were connected to the second question in the previous chapter, and, consequently, focussed on the significance of the difference in performance between the IKR and DKR treatment groups. The two treatments were expected to show a similar effect on performance in arithmetic.

The test of the simple effects, using the data on Occasion 1, was used to examine the similarity of the three treatment groups at the start of the experiment. The results indicated a significant difference, $F_{obs.} = 4.88$ (df 2, 699; $p < .01$). Consequently, comparisons were made on the gains in performance from Occasion 1 to Occasion 2, Occasion 2 to Occasion 3, and Occasion 1 to Occasion 3. Table 5.3 presents the gains in performance across the three occasions, while Table 5.4 displays a summary of the relevant Scheffe multiple comparisons tests of significance on these gains.

The first two rows of Table 5.4 indicate that the gains in performance obtained by the IKR and DKR treatment groups were very similar, the observed F-values being much smaller than the critical $F(2, 233) = 2.39$ at the .10 level of significance. This finding of no difference was in the direction expected.

The Main Effects of the Treatments

The last four comparisons in Table 5.4 contrast the gains of one

Table 5.3

Column Differences Between the Means

Treatments	Occasions			n
	2 - 1	3 - 2	3 - 1	
IKR	2.371	.342	2.617	81
DKR	1.843	.506	2.349	83
no-KR	.555	-.125	.430	72

Table 5.4

Scheffe Multiple Comparisons on the Differences Between the Gains of Each Treatment Group

Groups	Occasions	Diff.	$F_{obs.}$	Significance
IKR - DKR	2 - 1	.428	.957	N.S.
IKR - DKR	3 - 1	.278	.404	N.S.
IKR - no-KR	2 - 1	1.740	14.878	***
IKR - no-KR	3 - 1	2.187	23.335	***
DKR - no-KR	2 - 1	1.287	8.200	***
DKR - no-KR	3 - 1	1.919	18.201	***

Note. N.S. means not significant at the .10 level.

*** means significant beyond the .001 level.

treatment group with the corresponding gains of the control group. The third and fourth rows deal with the immediate posttest KR treatment, whereas the last two rows pertain to the delayed KR treatment group. In all instances, both treatment groups showed significant gains in performance, the observed F-values being larger than the critical $F(2, 233) = 6.91$ at the .001 level. These findings supported the prediction of a significant main effect of KR on performance in arithmetic. They were consistent with the test on the main effects displayed in Table 5.1. Also, when this test was done by contrasting the average gains taken over the two treatment groups with the corresponding gains for the control group, the differences in performance were, again, significant beyond the .001 level, suggesting that KR had a highly significant, positive effect on performance in arithmetic.

The Interaction of Academic Self Concept and the KR Treatments

This section was designed to measure certain aspects of cognitive mediation taking place as students interpret, react to, or use feedback. More specifically, it attempted to measure the covariation or influence of self concept of ability with the effectiveness of knowledge of results.

A total of nine one-way analyses of covariance were used in this section. In all of these, the three treatments served as the independent variable, with the performance in arithmetic scores at the three occasions (taken one at a time) as the criterion. The concomitant variables were the scores on the subscales of the Student's Perception of Ability Scale (Boersma and Chapman, 1977). Two important assumptions made with respect to these analyses state that the experimental groups were only randomly different on the covariates, and that the effects of the covariates on the criterion were linear, or nearly so.

The first section of the analyses aimed at identifying the significant covariates with the criterion, as well as those showing greatest change in predictability after the inception of the treatments. This involved the running of three one-way analyses of covariances (one on the scores taken at each occasion) with the six covariates considered together. In the second section, the analyses involved a separate use of each significant covariate in re-runs of this programme. The

purpose of this section was to assess the interaction of each signifi-cant concomitant with the effectiveness of the treatments.

Identifying the Most Useful Concomitants

While they identify the most significant covariates, the tests in this section bear some relevance to the question asked in Chapter III concerning the regression of self concept of ability in arithmetic on performance in the subject.

A summary of the one-way analysis of covariance test using Occasion 1 scores as the criterion, and scores on the six subscales as covariates, showed self concept of ability in arithmetic as the most significant concomitant. The observed F = 12.135, which was greater than $F(1, 227)$ = 10.83 at the .001 level of significance. Its positive regression estimate, (.443), suggests that its relationship with performance in arithmetic at this grade level was positive. Self concept of ability in reading was the second highest predictor with an F = 1.928, p = .166. All other covariates, including self concept of general academic ability, showed much lower estimates.

The analysis of covariance using Occasion 2 scores (taken one day after the first issue of KR) as the criterion and the scores on the six subscales as the covariates showed a significant change in the predictability of the reading self concept covariate. This concomitant was now significant, F = 4.958, which was greater than $F(1, 227)$ = 3.89

at the .05 level of significance. Self concept of ability in arithmetic remained significant with an observed $F = 5.904$, $p = .016$.

By Occasion 3, the covariates seemed to be moving back to their original positions, self concept of ability in reading showing an insignificant F of 2.439 ($p = .120$), while the values for self concept of ability in arithmetic were, $F = 7.997$, $p = .005$ (which was significant beyond the .01 level). These covariates, therefore, were included in the remaining analyses.

The findings in this section, then, were in the expected direction, suggesting that self concept of ability in arithmetic had a significant positive relationship with performance in the subject.

The Treatment X Self-Concept Interaction

As stated earlier, the question of interest here has to with the nature of the interaction between the significant self concept covariates and the effectiveness of the two types of feedback. This interaction was expected to be significant showing greater regression estimates for the IKR than for the DKR treatment group.

The analyses referred to here were from six one-way ANCOVA runs, using each of the post KR performance measures as criterion and the scores on each significant subscale as a covariate. The homogeneity of within cell regression estimates were of greatest relevance to these

FIGURE 5.2

INTERACTION OF SELF CONCEPT OF ABILITY IN ARITHMETIC WITH ARITHMETIC PERFORMANCE FEEDBACK AT OCCASION 2.

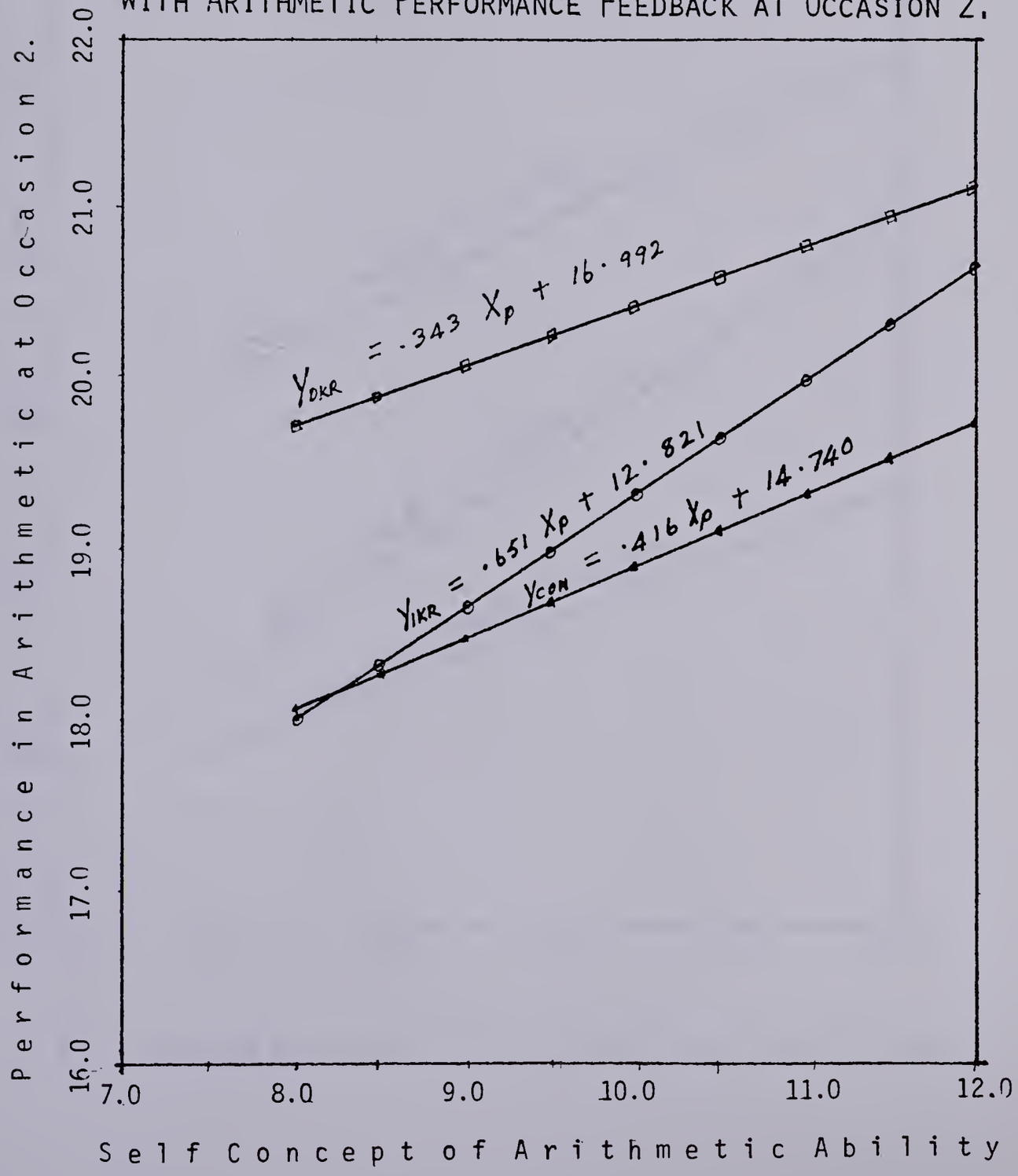


FIGURE 5.3

INTERACTION OF SELF CONCEPT OF ARITHMETIC ABILITY WITH ARITHMETIC PERFORMANCE FEEDBACK AT OCCASION 3.

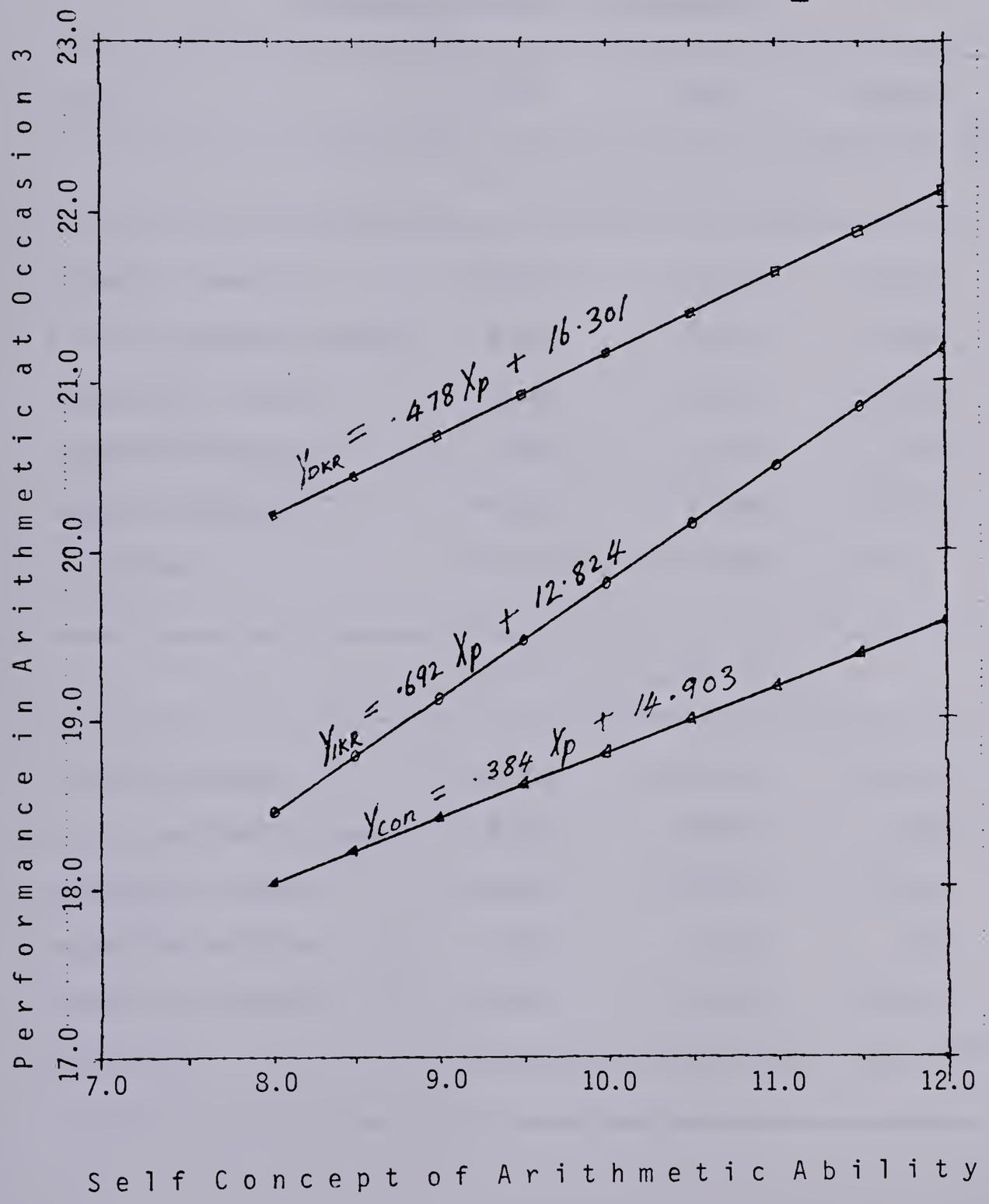


Table 5.5

Interaction of Self Concept of Ability in Arithmetic With Performance Feedback in Arithmetic

Source	IKR	DKR	Control
	O c c a s i o n 2		
Criterion (means)	18.222	20.036	18.486
S.C. in arithmetic (means)	8.296	8.880	9.014
Regression constant	12.821	16.992	14.740
Regression estimate	.651	.343	.416
Variance (adjusted)	26.534	15.198	26.438
F observed	df(2, 230)	F = .9699,	p = .381
	O c c a s i o n 3		
Criterion (means)	18.568	20.542	18.361
S.C. in arithmetic (means)	8.296	8.880	9.014
Regression constant	12.824	16.301	14.903
Regression estimate	.693	.478	.384
Variance (adjusted)	21.385	16.126	27.211
F observed	df(2, 230)	F = 1.033,	p = .358

considerations. These estimates are graphed in Figures 5.2 to 5.5. Other relevant details are displayed in Tables 5.5 and 5.6.

Table 5.5 indicates that the regression estimates were in the direction expected, the coefficients being slightly higher for the IKR than for the DKR treatment. The values observed after one issue of KR were .6511 and .3428 for the immediate posttest KR and the delayed KR treatments, respectively. They showed little change after the second issue of the treatment.

The differences in regression, however, were not significant. The tests over all three treatment groups are displayed in Table 5.5. For the two KR treatments only, the observed $F(1, 160) = 1.831$ ($p = .178$). at Occasion 2; and for Occasion 3, $F(1, 160) = .986$ ($p = .322$). The positive slopes, or regression estimates, (see Figures 5.2 and 5.3) indicate positive influences of the covariate on both treatments, suggesting that higher self concept students did better than lower self concept students under both immediate posttest KR and one-day dealyed KR treatments. The delayed KR was the more effective treatment for students in both self concept categories.

Differences between the regression estimates were larger for the self concept of ability in reading concomitant. At Occasion 2, the estimates observed were, .794 and .197 for the IKR and the DKR treatment groups, respectively; and at Occasion 3, the corresponding regression estimates were .849 and .143. Table 5.6 indicates that the differences

FIGURE 5.4

INTERACTION OF SELF CONCEPT OF READING ABILITY WITH ARITHMETIC PERFORMANCE FEEDBACK AT OCCASION 2.

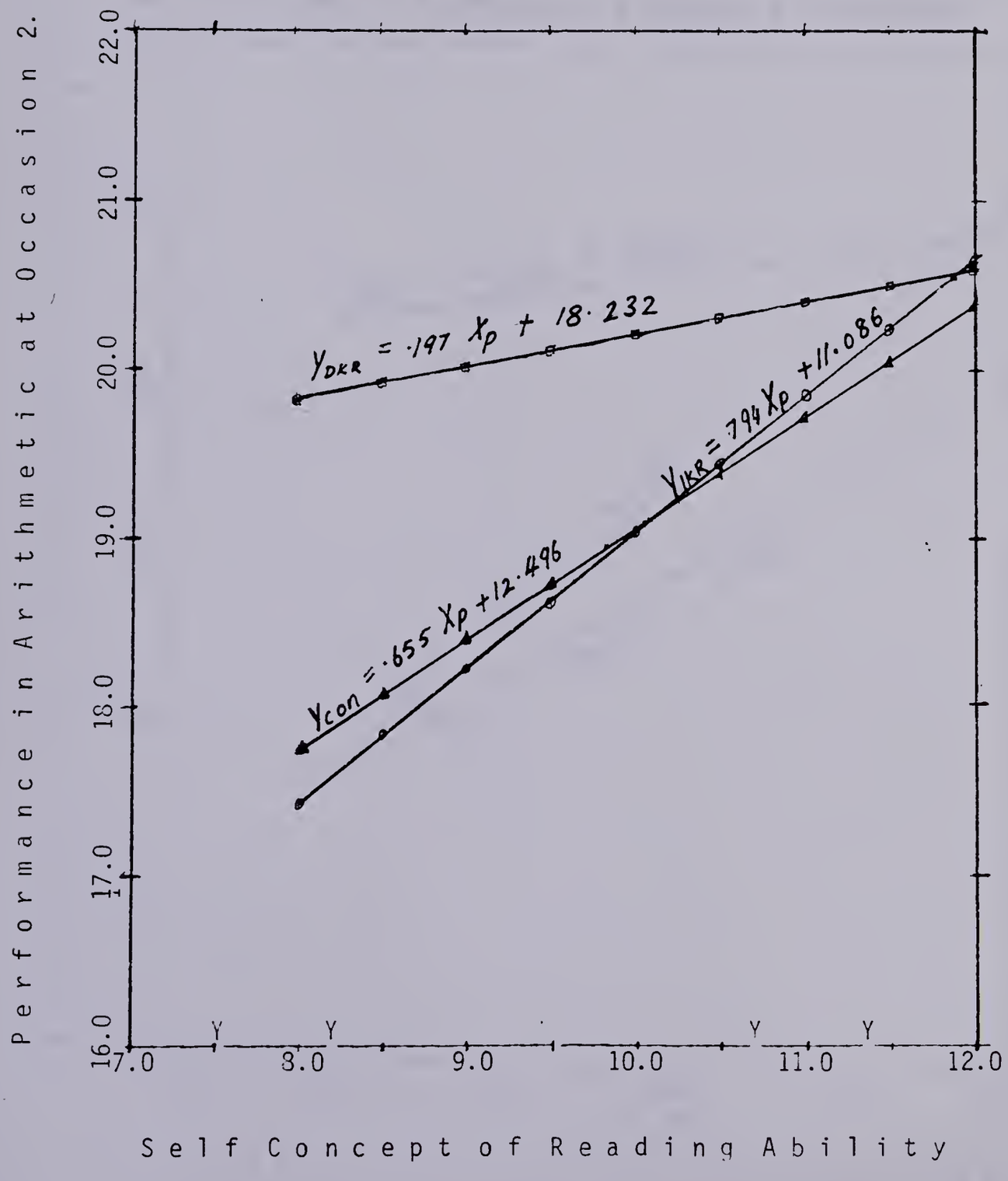


FIGURE 5.5

INTERACTION OF SELF CONCEPT OF READING ABILITY WITH ARITHMETIC PERFORMANCE FEEDBACK AT OCCASION 3

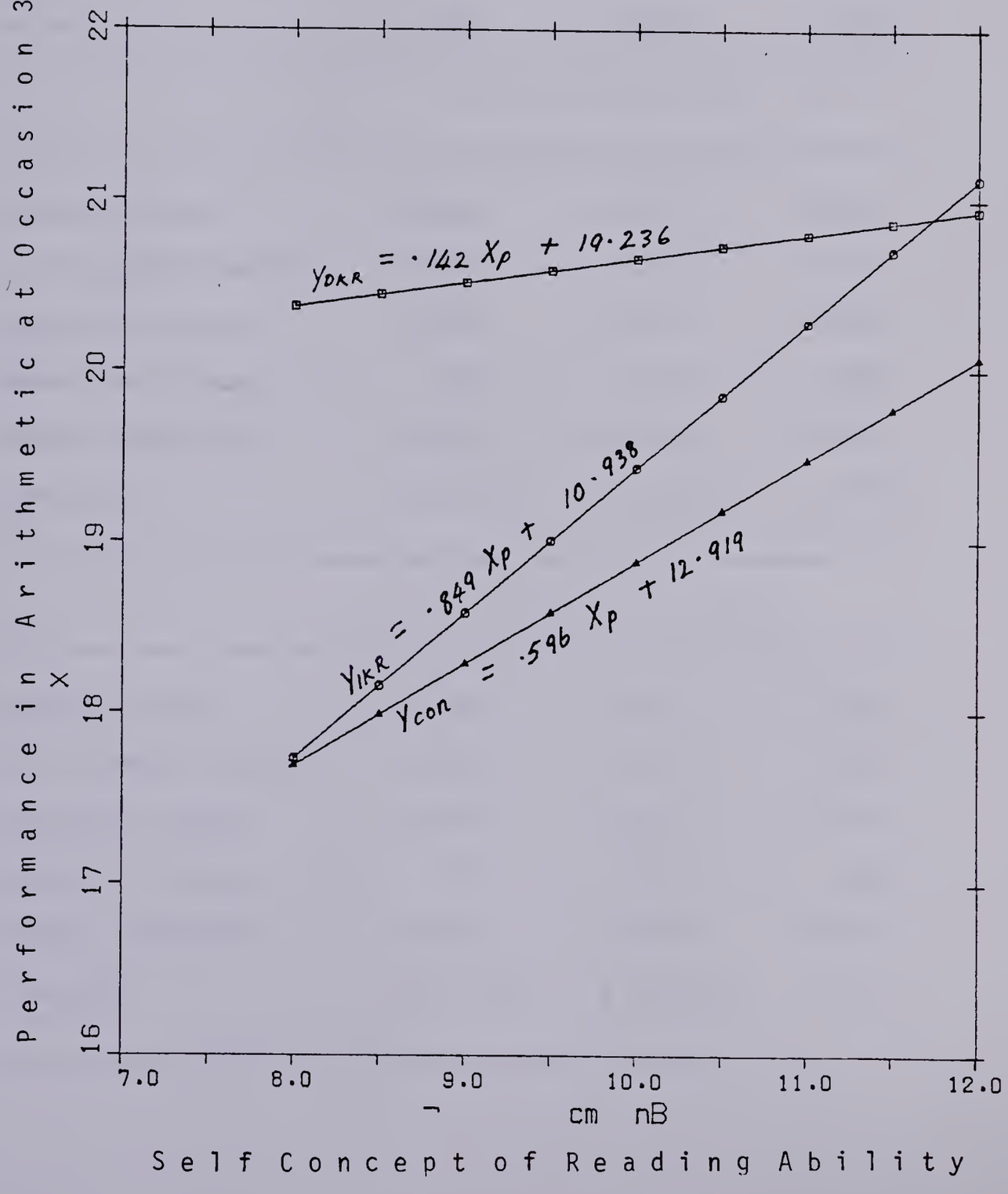


Table 5.6

Interaction of Self Concept of Ability in Reading With Performance Feedback in Arithmetic

Source	IKR	DKR	Control
	O c c a s i o n 2		
Criterion (means)	18.222	20.036	18.486
S.C. in reading (means)	8.988	9.169	9.139
Regression constant	11.086	18.232	12.496
Regression estimate	.794	.197	.656
Variance (adjusted)	26.455	15.934	23.625
F observed	df(2, 230)	F = 2.839,	p = .0606
	O c c a s i o n 3		
Criterion (means)	18.568	20.542	18.361
S.C. in reading (means)	8.988	9.169	9.139
Regression constant	10.938	19.236	12.919
Regression estimate	.849	.143	.596
Variance (adjusted)	21.234	17.974	24.944
F observed	df(2, 230)	F = 3.800,	p = .024

for the three groups considered together reached statistical significance by Occasion 3. When the regression estimates for the two KR treatments were compared separately, the differences between them were seen to be significant on both post-feedback occasions, the observed values being: $\underline{F}(1, 160) = 5.286$ ($\underline{p} = .023$) at Occasion 2, and $\underline{F}(1, 160) = 7.981$ ($\underline{p} = .005$) at Occasion 3. It should be noted that the three regression estimates were very similar at Occasion 1, before the first issue of the treatment, the observed $\underline{F} = .496$ which was less than the critical $_{.95}\underline{F}(2, 230) = 3.04$.

The interaction between self concept of ability in reading and performance in arithmetic was disordinal suggesting that immediate posttest KR might be better for students at the uppermost reading self-concept level, whereas one-day delayed KR was otherwise the better treatment. The findings with respect to the proposition of a significant interaction between the effectiveness of the KR treatments and the significant self concept concomitants, therefore, were in the direction expected. The higher significance of the interaction with respect to self concept of ability in reading supports earlier conclusions (see Cronbach and Snow, 1977) that reading is the greatest task of the student in his first years at school.

Preference for IKR or DKR

This section of the analyses dealt with the distribution of the students' preferences over the three options: feedback immediately after, one day after, or one week after finishing the test. Concerns about the interaction of the students' preferences with the effectiveness of the KR treatments, and the influences of these preferences on the powers of the treatments to produce the desired results were secondary since the students were asked to think about their preferences after the treatments were issued. These latter effects, therefore, were expected to be insignificant.

Table 5.7

Frequencies of the Students' Preferences for Type of Feedback

Treatments	Feedback preferences		
	IKR	1-day DKR	1-week DKR
IKR	41	28	12
DKR	37	37	9
no-KR	28	29	15
Totals	106	94	36

categories is displayed in Table 5.7. The chi square test on the total sample, taking all three categories, indicated significance beyond the .001 level, $\chi^2_{obs.} = 35.627 > _{.999}\chi^2_{(2)} = 13.82$. The distribution of the preferences of this sample of students, then, suggests a strong tendency for students in this population to prefer shorter than longer delays of performance feedback. A close look at Table 5.7 shows that much of the variance tapped by the chi square value mentioned above was attributable to the fact that few students preferred to know their results one week after completing the test. A second, but much smaller source of variance came from the fact that as much as 50.5 per cent of the students treated with immediate feedback preferred this treatment, whereas the other 49.5 per cent was distributed over the other two preference categories. The distributions for the delayed KR and the control groups were quite similar across the IKR and the 1-day DKR preference categories.

The second test investigated the student's preference for immediate posttest KR versus one-day delayed KR, and consequently, combined the previously used delay categories into one. It was assumed here that students who preferred to have their feedback one week after finishing the test would have chosen to have their results one day after, had they been given only the first two options - viz feedback immediately after, or one day after completing the test -; and that the students who had indicated a preference for one-day delayed KR preferred this option over knowing their results immediately, and consequently, would have remained

in this category had they been given the first two options only. Obviously, this assumption ignores the operation of the regression toward-the-mean phenomenon on the influence of the students' choices in the first situation where they were given three options.

The observed proportions were 106/236 for IKR and 130/236 for DKR. Though it seems as if there was a slight tendency to favour one-day delayed feedback over immediate posttest feedback, the difference between these two proportions was not significant, $\chi^2_{obs.} = 2.44 < {}_{.95}\chi^2_{(1)} = 3.84$.

The preference X treatment and the preference effects. The analyses referred to in this section were obtained from three two-way analyses of covariance. Each of the performance in arithmetic scores served as a criterion with the KR treatments as an independent variable. The students' preferences were used as a second independent variable (which may be better referred to as a stratification variable), and both significant concomitants (self concept of ability in arithmetic, and self concept of ability in reading) served as covariates.

The lack of previous research in this area prevented the making of any prediction about the treatment X preference interaction. It was thought, however, that since the students were asked to make a choice after they had been treated, this interaction would be insignificant. Table 5.8 displays the relevant F-values and shows that this interaction was insignificant at all three occasions. The students' preferences

Table 5.8

A Summary of the Treatment X Preference Interaction Effects

SS	SS	df	MS	$F_{obs.}$	Prob.	Sig.
Occasion 1						
	90.9695	4, 225	22.7424	1.0766	.3688	N.S.
Occasion 2						
	61.7804	4, 225	14.9839	.7077	.5874	N.S.
Occasion 3						
	101.6066	4, 225	25.4017	1.2222	.3022	N.S.

showed a significant effect on their performance in arithmetic before the inception of the treatments, $F(2, 225) = 3.469$; $p = .033$. However, this effect was insignificant on measures of the criterion taken after the students were treated with feedback, $F(2, 225) = .6866$ ($p > .05$) at Occasion 2, and $F(2, 225) = .5852$ ($p > .05$) at Occasion 3.

A Summary of the Findings

The analyses indicated findings which supported, very decisively, the major expectations of the study. With respect to the influence of KR on performance in arithmetic, the level of statistical significance in every instance was beyond .001.

The two types of feedback treatments (immediate posttest KR and one-day delayed KR) seemed from the tests of analysis of variance with repeated measures, to be quite similar with both treatment groups showing significant performance gains.

Self concept of ability in arithmetic was, in every instance, a clearly significant concomitant with performance in the subject, although its effectiveness showed a slight drop after the students were treated with KR. Self concept of ability in reading, on the other hand, emerged from a position of insignificance before the first issue of KR to one of significance beyond the .01 level after only one issue of the treatment. By Occasion 3, one week after the second issue of the treatments, the concomitant had returned to its former position of insignificance. The finding of significantly positive regression estimates of self concept of ability in arithmetic supported the proposition of a positive relationship between the effectiveness of feedback and this self concept measure.

The interaction of self concept of ability in arithmetic with the feedback treatments was not significant. The estimates indicated a positive influence on the effectiveness of both KR treatments, but the difference between these estimates never approached any level of statistical significance. Self concept of ability in reading, on the other hand, interacted significantly with the effectiveness of the two KR treatments with the difference in regression estimates reaching the .01 level of significance by Occasion 3 (see Figures 5.4 and 5.5).

The students showed a strong tendency to prefer shorter than longer delays of feedback. The tendency was most pronounced for students treated with IKR, the majority of whom preferred to know their results immediately after completing the test. However the difference between the entire sample's preference for immediate posttest KR or one-day delayed KR was not significant.

VI

Discussion and Limitations

The finding that feedback influenced performance in arithmetic was quite consistent with the theoretical notion that feedback affects performance, as well as with the large body of research on performance feedback. Few, if any, of the previous research efforts had examined the effectiveness of feedback on performance in this discipline, with students of this age level. White (1968) and O'Neil, Rasor and Bartz (1979) came closest. O'Neil and his associates used a 50-item, objective type IQ test of general ability, and thus included a few arithmetic questions. Their subjects, however, were 116 volunteers from under graduate psychology classes. It is interesting to note that both studies reported a similar level of significance ($p<.001$) with the effectiveness of feedback on performance. In their study, as it was in this one, the no-feedback group did not increase its level of performance, whereas all groups treated with feedback made highly significant performance gains irrespective of the type of feedback they received.

The finding that immediate posttest knowledge of results was not better than knowledge of results delayed for one day may be seen as consistent with the bulk of the research in the related literature, as well as with the theory of response competition or interference perseveration. To appreciate this high degree of consistency, however, one

would have to bear in mind the definition of 'Immediate feedback' as used in this study. In the O'Neil et al (1979) study, the performance of the immediate posttest feedback group was clearly the highest, being significantly higher ($p<.05$) than those of the two more immediate feedback treatment groups. Their study did not include feedback issued after longer delays. On the premise of this finding that immediate post-test feedback is superior to immediate item-per-item feedback, as well as on the assertion of the interference-perseveration hypothesis, one would expect the influences of immediate posttest feedback and one-day delayed feedback to be more similar than the influences of immediate item-per-item feedback and one-day delayed feedback.

Several, if not the majority, of the studies reporting a significant difference between the effectiveness of immediate feedback after each item and delayed feedback noted significance at only the .05 level. Some of these studies reported a significant difference only after the use of a very sensitive criterion measure (R_2/W_1) referred to as the ratio of the subjects changing their wrong responses on Occasion 1 (W_1) to the right responses on Occasion 2 (R_2). It is not alarming then, that from a less sensitive measure (the number right score) the difference between the effectiveness of immediate posttest KR and one-day delayed KR was not significant.

The number right score was used in this study since it is the measure used by the class teacher, and since the use of parallel forms of the arithmetic achievement test, rather than the same form on all three

occasions, made the use of R_2/W_1 impossible or at least dubious. Also, the use of R_2/W_1 assumes that the extent to which feedback on one item affects performance on another item is neglible. Though this assumption may not be far from the truth in retention and recall of verbal material this may not be the case in tests of arithmetic operations. In the latter type of test items, each sum is a combination of several operations which may overlap from one problem to another. For example, the large majority of the items in each of the three test forms used to measure arithmetic achievement involved the use of addition facts and the regrouping of numbers. A check on the students' responses across the three occasions indicated that getting item 'i' right on Occasion 1 was not a sufficient condition for doing its parallel correctly on Occasion 2 and Occasion 3.

When the results of this study were compared with previous findings, the question was raised as to how much time should elapse between the response and the feedback in order to promote maximum learning. O'Neil et al's (1979) immediate posttest feedback was associated with greater learning than was immediate item-per-item feedback; the present study suggested that immediate posttest feedback was as effective as one-day delayed feedback; and White's (1968) immediate posttest feedback treatment was superior to feedback delayed for three days. This trend suggests a feedback effectiveness curve which is at its maximum point somewhere between immediately after and one day after completing the test. This suggestion supports More's (1969) conclusion that the maximum point is somewhere between $2\frac{1}{2}$ hours and 24 hours after completing the test.

It should be noted that the author disagrees with the assumption of White (1968) that cheating was distributed evenly across his treatment groups. Subjects given the correct answers before handing in their responses to the investigator were more likely to cheat than others given the correct answers after the investigator had taken home their answer sheets for the weekend. Indeed cheating may have been at its highest in White's immediate item-per-item feedback situation. In the present study, cheating among both KR treatment groups was likely to be very low since i) the electronic scorer printed the scores and other marks in indelible ink; ii) students were given a chance to better their scores via the rescore option explained in Chapter IV (though the second and third scores were not considered in the analyses); and iii) the students were told that the grades obtained on these tests would not go on their report cards, that the exercises were designed to give them practice on some useful arithmetic skills, and consequently, their improvement or second score was as important as their first score.

The finding that self concept of ability in arithmetic was significantly related to performance in arithmetic was consistent with the correlational research in this area. The observed regression estimates suggested a relationship which was very similar to Brookover et al's (1964, 1967) measures taken independently of the influence of IQ. Work relating self concept measures taken on the Student's Perception of Ability Scale (Boersma and Chapman, 1977) with IQ or actual ability measures is being undertaken at present. However, data collected by Chapman and Boersma (1979), using the Otis Lemon Form K for Grades 3

and 4, and Lorge-Thorndike for Grades 5 and 6, indicated findings which were consistent with the theoretical notion that successful and unsuccessful students developed different self perceptions of their academic ability independent of actual ability (Bloom, 1976). The regression estimates observed in the present study, therefore, may be suggesting a genuine relationship between self concept of ability in arithmetic and performance in the subject. That the covariation was significant at every stage of the experiment suggests that this relationship is reliable.

The finding that at least one self concept measure (self concept of ability in reading) interacted significantly with the effectiveness of the treatments lends some support to the view that some sort of processing or interpretation of the conjoint information is involved as the student accepts and responds to performance feedback, and consequently, to the hypothesis of cognitive mediation in the learning process. Performance gains, then, may not result purely from automatic reactions to positive stimuli, though the findings of this study do not rule out the possibility that there is some influence of this nature.

That the regression estimates with respect to the IKR treatment were higher than those associated with the DKR treatment (the regression estimates for the self concept of ability in arithmetic concomitant showed this trend, also) supports the theory that interference or response competition operates in immediate feedback situations. It suggests that whereas high self concept students in an immediate feedback

situation may overcome inhibitions in the process of learning by their more positive attitude and possibly greater effort, lower self concept students seem not to overcome these inhibitors as easily, and consequently, do less well than other low self concept students in an interference free situation. Though the initial performance of DKR subjects was higher than that of IKR subjects, the trend observed seemed not to be the result of a ceiling affecting the former group more than the latter since the means and standard deviations suggested enough room to accomodate increases in performance, and the overall gains for the two treatment groups were very similar (see Tables 4.1, 5.2, 5.4, 5.5 and 5.6)

The finding that self concept of ability in reading was as reactive, or unstable, as self concept of ability in arithmetic throughout the treatment phase of the experiment was interesting. The regression slopes were: .406 ($p = .0004$), .278 ($p = .0163$), and .343 ($p = .0025$) for self concept of ability in arithmetic; and .244 ($p = .0504$), .353 ($p = .0055$), and .307 ($p = .0131$) for self concept of ability in reading at Occasions 1, 2 and 3, respectively. This comparison of the regression estimates shows self concept of ability in reading as escalating from a position of insignificance at Occasion 1 to one of significance beyond the .01 level on scores taken after the first issue of the treatment. Self concept of ability in arithmetic, on the other hand, was reduced from significance beyond the .001 level on Occasion 1 to a lower significance which was beyond only the .05 level. Though this behaviour of the covariates supports the claims of Bilodeau and Bilodeau (1958), and Tait, Hartley and Anderson (1973) that retention is largely a verbal phenomenon, and the conclusion that reading is the greatest task of the student in his first years at

school (Cronbach and Snow, 1977), one would expect a higher interaction between self concept of ability in arithmetic and the effectiveness of performance feedback in arithmetic, than between these treatments and self concept of ability in reading. The finding that greater transactions take place with the latter construct, however, gives added prominence to the view that the effectiveness of feedback depends in part on the subject's ability to interpret and use the conjoint information.

The findings with respect to the students' preferences clearly suggest that this population of students prefers shorter than longer delays of feedback. Also, it should be noted that students who had experienced the immediate feedback treatment showed greatest preference for this type of feedback. Granted that KR has some attributes of a reward, this finding is consistent with the research on animals (Davenport, 1962; Logan, 1968; Ainslie, 1974) that the immediate reward is preferred over the delayed reward. The feedback stimulus, however, may have characteristics other than those common to rewards since it is accompanied by information which may be positive or negative, confusing or enlightening when it is related to previous knowledge and experience. The finding of no difference between the students' preference for immediate posttest KR and one-day delayed KR adds some relevance to this point. This latter finding may be the result of other factors, like degree of confusion, of the feedback situation, and consequently, may be indicating that, other things being equal, students show less liking for a confusing situation than for one that is void of conflict.

A Statement of Limitations

Though considerable effort was made to obtain items between the .4 and .6 difficulty levels, a somewhat large number of the items on the final forms were easy, falling between the indices of .6 and .8. Though there was room to accomodate increases in performance, more difficult items would have increased the error rate, and, consequently, the amount of interference or conflict between the correct and incorrect responses. In the same vein, items of slightly higher discrimination indices could have contributed to the major task of discriminating between the performance of subjects in the IKR and DKR treatment groups.

The task of preparing three forms of an instrument that is just right, however, takes a little more time than could be afforded in this study. In replications of these findings, if the researcher wishes to focus on the arithmetic operations of addition and subtraction, it may be easier for him to obtain items of a high difficulty level with Grade 3 than with Grade 4 students.

Previous hints have been made at the use of a more sensitive measure of the criterion. Though a closer examination suggests that the use of R_2/W_1 as the criterion would have resulted in similar findings in this study, it is most often worthwhile to give some thought to the use of a more sensitive criterion.

That there were initial differences between the treatment groups was unfortunate. This resulted primarily from the fact that the overall

sample was distributed over the treatments in clusters, and not on a purely random basis. Though this method produced an undesireable situation here, the alternative was likely to result in confusion in each classromm and greater influence of extraneous variables, and, consequently should not be preferred in experiments of this nature.

Suggestions for Further Research

It would be interesting to replicate this study using more difficult items in the arithmetic tests so as to increase the likelihood of conflict between the correct and incorrect responses. The findings with respect to the interactions between the significant self concept concomitants and the effectiveness of the treatments will do well with further investigation. It may be that the very strong relationship between self concept of ability in arithmetic and performance in the subject made it a significant covariate in all feedback situations, and consequently, a poor discriminator between the effectiveness of the treatments. However, this explanation needs to be reexamined in a subsequent research.

Further studies connecting preference for type of feedback and feedback effectiveness are needed if one is to knew how this aspect of the research compares with the propositions of the theory of interference-perseveration. Subsequent research efforts in this area may do better if they were to focus on the degree of conflicting information presented with the feedback and reduce the conjoint rewarding elements of the construct.

REFERENCES

AINSLIE, G. Impulse control in pigeons. Journal of Experimental Analysis of Behavior, 1974, 21, 485-489.

AINSLIE, G. Specious reward: A behavioral theory of impulsiveness and impulse control. Psychological Bulletin, 1975, 82, 463-496.

ALEXANDER, L.T., Knowledge of results and the temporal gradient of reinforcement. American Psychologist, 1951, 6, 292-293.

AMMONS, R.B., Effects of knowledge of performance: A survey and tentative theoretical formation. Journal of Central Psychology, 1956, 55, 279-299.

ANDERSON, R.C., KULHAVY, R.W. and ANDRE, T. Feedback procedures in programmed instruction. Journal of Ed. Psychol., 1971, 62, 148-156.

ANNETT, J. Feedback and human behavior. Great Britain: Penguin Books, 1969.

BACKMAN, C.W., SECORD, P.F. A social psychological view of education. New York: Harcourt, Brace & World, 1968.

BECK, F.W., LINDSEY, J.D. Effects of immediate information feedback and delayed information feedback on delayed retention. Journal of Educational Research, 1979, 72, 283-287.

BECKER, J.L. The effects of withholding reinforcement in auto-instructional programmes, In Ofiesh, G.D., & Meierhenry, W.C. Trends in programmed instruction. Washington: Davi N s p i, 1964.

BECKER, J.L., MUSSINA, C.M., PERSONS, R.W. Intertrial interview, delay of knowledge of results, and motor performance. Perception of Motor Skills, 1963, 17, 559-563.

BILODEAU, E.A. A further study of the effects of target size and goal attainment upon the development of response accuracy. U S A F H R R C, Research Bulletin, 1952, 52-57.

BILODEAU, E.A. Speed of acquiring a simple motor response as a function of the systematic transformation of KR. American Journal of Psychology, 1953, 66, 409-420.

BILODEAU, E.A. Retention, In Bilodeau, E.A. (Ed.) Acquisition of Skill. New York: Academic Press, 1966, pp. 315-350.

BILODEAU, E.A., BILODEAU, I.McD. Variable frequency of knowledge of results and the learning of a simple skill. Journal of Experimental Psychology, 1958, 55, 379-383.

BILODEAU, E.A., ROSENBACH, J.H. Acquisition of response proficiency as

a function of rounding error in informative feedback. H R R C, Research Bulletin, 1953, 53-121.

BILODEAU, E.A., RYAN, F.J. A test for interaction of delay of knowledge of results and two types of interpolated activity. Journal of Experimental Psychology, 1960, 59, 414-419.

BLACK, F.W. Self concept as related to achievement and age in learning disabled children. Child Development, 1974, 45, 1137-1140.

BOERSMA, F.J., CHAPMAN, J.W., BATTLE, J. Academic self-concept in special education students: Some suggestion for interpreting self-concept scores. The Journal of Special Education, 1979, 13, 433-441.

BLOOM, R.B., BOURDOM, Linda. Types and frequencies of Teacher's written instructional feedback. Journal of Educational Research, 1980, 74, 13-18.

BOERSMA, F.J., CHAPMAN, J.W. The Student's Perception of Ability Scale. Unpublished instrument, University of Alberta, 1977.

BOERSMA, F.J., CHAPMAN, J.W., MAGUIRE, T.O. The Student's Perception of Ability Scale: An instrument for measuring academic self-concept in elementary school children. Educational and Psychological Measurement, 1979, 39, 1035-1041.

BOURNE, L.E. Human conceptual behavior. New York: Allyn and Bacon Inc. 1966a.

BOURNE, L.E. Information feedback: Comments on Professor I.McD. Bilodeau's paper. In Bilodeau, E.A. (Ed.), Acquisition of Skill. New York: Academic Press, 1966, pp. 297-313.

BOURNE, L.E., BUNDERSON, C.V. Effects of delay of information feedback and length of post feedback interval on concept identification. Journal of Experimental Psychology, 1963, 65, 1-5.

BOURNE, L.E., GUY, D.E., WADSWORTH, N. Verbal reinforcement combinations and the relative frequency of informative feedback in a card sorting task. Journal of Experimental Psychology, 1967, 73, 220-226.

BROOKOVER, W.B., ERICKSON, E.L., JOINER, L.M. Self concept of ability and school achievement, III: Relationship of self concept to achievement in high school (Educational Research Series No. 36). East Lancing, Mich.: Educational Publication Series, 1967.

BROOKOVER, W.B., LePERE, J.M., HAMACHEK, D.E., THOMAS, S., ERICKSON, E.L. Self concept of ability in school achievement, II (Educational Research Series No. 31). East Lancing, Mich.: Educational Publication Series, 1965.

BROOKOVER, W.B., THOMAS, S., PATERSON, A. Self concept of ability and school achievement. Sociology of Education, 1964, 37, 271-278.

BUCHWALD, A. M. Effects of immediate vs delayed outcomes in associative learning. Journal of Learning and Verbal Behaviour, 1967, 6, 317-320.

BUCHWALD, A.M. Effects of right and wrong on subsequent behaviour: A new interpretation. Psychological Review, 1969, 76,132-142.

BUCHWALD, A. M., MEAGER, R. B. Immediate and delayed outcomes: Learning and recall of responses. Journal of Experimental Psychology, 1974, 103, 758-767.

BUSS, A. H., BRADEN, W., ORGEL, A., BUSS, E. H. Acquisition and extinction with different reinforcement combinations. Journal of Experimental Psychology, 1956, 52, 288-295.

BUSS, A. H., BUSS, E. H. The effects of verbal reinforcement combinations on conceptual learnong. Journal of Experimental Psychology, 1956, 52, 283-287.

CAMPBELL, D. T., STANLEY, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally Co., 1963.

CASLYN, R. J., KENNY, D. A. Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement. Journal of Educational Psychology, 1977, 63, 136-145.

CHAPANIS, A. Knowledge of performance as an incentive in repetitive tasks. Journal of Applied Psychology, 1964, 48, 263-267.

CHAPMAN, J. W., BOERSMA, F. J. Academic self-concept in elementary learning disabled children: A study with the Student's Perception of Ability Scale. Psychology in Schools, 1979, 16, 201-206.

COPPERSMITH, S. The antecedents of self-esteem. San Francisco: W. H. Freeman & Co., 1967.

COWLES, J. T. Food tokens as incentives for learning by chimpanzees. Comp. Psychol. Monogr., 1937, 14, #71.

CRONBACH, L. J., SNOW, R. E. Aptitudes and instructional methods; A handbook for research in interactions. New York: Irvington Publishers Inc., 1977.

D'AMATO, M. R., COX, J. K., Delay of consequences and short-term memory

in monkeys. In D. L. Medin, W. A. Roberts, & R. T. Davis (Eds.) Processes in animal memory. Hillsdale, New Jersey: Erlbaum Associates, 1976.

DANNEMBERG, R. A. The relationship of intelligence to reinforcement in a linear programmed instruction unit dealing with the free enterprise system. University of Michigan: Dissertation Abstracts, 1965, 26, 260-268.

DAVENPORT, J. The interaction of magnitude and delay of reinforcement in spatial discrimination. Journal of Comparative and Physiological Psychology, 1962, 55, 267-273.

DECI, E. L. Intrinsic motivation. New York: Plenum Press, 1975.

EPPS, E. G. Correlates of academic achievement among northern and southern urban negro students. Journal of Social Issues, 1969, 35, 55-70.

FELDHUSEN, J. F., BIRT, A. A study of nine methods of presentation of programmed learning material. Journal of Educational Research, 1962, 55, 461-466.

FITTS, P. M. Factors in complex skill training. In R. Glaser (Ed.), Training Research and Education. University of Pittsburg Press: Wiley, 1962, rep. 1965.

GORDON, N. B. Personal communocation. New York: Academic Press, 1968.

GRELLER, M. M., HEROLD, P. M. Sources of feedback: A preliminary investigation. Organizational Behavior and Human Performance, 1975, 13, 244-256.

GREENSPOON, J., FOREMAN, S. Effects of delay of knowledge of results on learning a motor task. Journal of Experimental Psychology, 1956, 51, 226-228.

HACKMAN, J. R., OLDHAM, G. R. Development of the job diagnostic survey. Journal of Applied Psychology, 1975, 60, 159-170.

HACKMAN, J. R., OLDHAM, G. R. Motivation through the design of work: Test of a theory. Organizational Behavior and Human Performance, 1976, 16, 250-279.

HALPERIN, K., SNYDER, C. R., SHENKEL, R. J., HOUSTON, B. K. Effect of source status and message favorability on acceptance of personality feedback. Journal of Applied Psychology, 1976, 61, 85-88.

HAMMOND, K. R., SUMMERS, D. A. Cognitive control. Psychological Review, 1972, 79, 58-67.

HEROLD, D. M., GRELIER, M. M. Feedback: The definition of a construct. Academy of Management Journal, 1977, 20, 142-147.

HOLDING, D. H. Guidance in pursuit tracking. Journal of Experimental Psychology, 1959, 47, 374-382.

HOUSE, R. E., GLASS, G. V., McLEAN, L. D. WALKER, D. F. No simple answer: Critique of Follow Through evaluation. Howard Educational Review, 1978, 48,

HULL, C. L. Essentials of behavior. New Haven: Yale University Press, 1951.

ILGEN, D. R., HAMSTRA, B. W. Performance satisfaction as a function of the difference between expected and reported performance at five levels of reported performance. Organizational Behavior and Human Performance, 1972, 7, 359-370.

ILGEN, D. R., FISHER, C. D., TAYLOR, M. S. Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 1979, 64, 348-373.

JACOBS, M., JACOBS, A. FELDMAN, G., CAVIOR, N. Feedback II - The credibility gap: Delivery of positive and negative and emotional and behavioral feedback in groups. Journal of Consulting and Clinical Psychology, 1973, 41, 215-223.

JACOBS, P. I., KULHARNI, S. A test of some assumptions underlying programmed instruction. Psychological Report, 1966, 18, 103-110.

JAMES, W. H., ROTTER, J. B. Partial and 100 per cent reinforcement under chance and skill conditions. Journal of Experimental Psychology, 1958, 397-403.

KANFER, F. H., KAROLY, P. NEWMAN, A. Source of feedback, observational and attitude change. Journal of Personality and Social Psychology, 1974, 29, 30-38.

KANTOWITZ, B. H. (Ed.) Human information processing: Tutorials in performance and cognotion. Hillsdale, N. J.: Erlbaum, 1974.

KULHAVY, R. W., ANDERSON, R. C. Delay-retention effect with multiple choice tests. Journal of Educational Psychology, 1972, 63, 505-512.

KIPPEL, G. M. Information feedback schedules, interpolated activities and retention. Journal of Psychology, 1974, 87, 245-251.

KORMAN, A. K. Toward a hypothesis of work behavior. Journal of Applied Psychology, 1970, 54, 31-41.

KORMAN, A. K. A hypothesis of work behavior revisited and an extension. Academy of Management Review, 1976, 1, 50-63.

LEVINE, M., LEITENBERG, H., RICHTER, M. The blank trials law: the equivalence of positive reinforcement and non-reinforcement. Psychological Review, 71, 94-103.

LOCKE, E. A. The relationship of intentions to level of performance. Journal of Applied Psychology, 1966, 50, 60-66.

LOCKE, E. A. Motivation effects of knowledge of results on goal setting. Journal of Applied Psychology, 1967, 51, 324-329.

LOCKE, E. A., BRYAN, J. F. Cognotive aspects of psychomotor performance: the effects of performance goals on level of performance. Journal of Applied Psychology, 50, 286-291.

LOCKE, E. A., BRYAN, J. F. The effects of goal setting, rule learning and knowledge of score on performance. American Journal of Psychology, 79, 451-457.

LOGAN, F. A. Decision making by rats: Delay versus amount of reward. Journal of Comparative and Physiological Psychology, 1965, 59, 1-12.

LORGE, I., THORNDIKE, E. L. The influence of delay in the after effect of a connection. Journal of Experimental Psychology, 1935, 18, 186-194.

MACE, C. A. Incentives: some experimental studies. Industrial Health Research Board Report, #72, 1935.

MARX, M. H., GOODSON, F. E. Further gradients of error reinforcement following repeated reinforced responses. Journal of Experimental Psychology, 1956, 51, 421-428.

McGUINGAN, E. J. The effect of delay, precision, and schedule of knowledge of results on performance. Journal of Experimental Psychology, 1959, 58, 79-84.

MEYER, S. R. Report on the initial test of a junior high school vocabulary program. In A. A. Lumsdaine, & R. Glaser (Eds.) Teaching machines and programmed learning. Washington: NEA-NRC, 1960, pp. 229-246.

MEYER, H. H., WALKER, W. B. A study of factors relating to the effectiveness of a performance appraisal program. Personnel Psychology, 1961, 14, 291-298.

MICHAEL, D. N., MACCOBY, N. Factors influencing the effects of student participation on verbal learning from film: Motivating vs practice effects, "feedback" and overt vs covert responding, In A. A. Lumsdaine (Ed.), Student response programmed instruction, Washington: NAS-NRC, 1961.

MOORE, J. W., HALPERIN, J. Three-stimulus two-choice auditory discrimination learning with blank trials. Journal of Experimental Psychology, 1967, 73, 241-246.

MOORE, J. W., SMITH, W. J. Knowledge of results in self-teaching spelling. Psychological Report, 1961, 9, 717-726.

MORE, A. J. Delay of feedback and the acquisition and retention of verbal materials in the classroom. Journal of Educational Psychology, 1969, 60, 339-342.

O'NEIL, M., RASOR, R. A., BARTZ, W. R. Immediate retention of objective test answers as a function of feedback complexity. Journal of Educational Research, 1979, 72, 72-75.

OPPENHEIM, N. The relation between intelligence and difference patterns of feedback in a linear teaching program. Columbia University Dissertation Abstract, 1964, 25, 2857.

PARKINSON, B. F. The effects of an extended post information feedback interval, anxiety and ability on programmed learning with college students. Univ. Utah. Dissert. Abstr., 1964, 25 1753-1754.

PAYNE, R. B., HAULY, G. T. Effects of psychological feedback upon work decrement. Journal of Experimental Psychology, 1955, 50, 342-351.

PEECK, J., TILLEMA, H. H. Delay of feedback and Retention of correct and incorrect responses. Journal of Experimental Education, 1979, __, 171-178.

PHARES, E. J. Expectancy changes in skill and change situations. Journal of Abnormal Social Psychology, 1957, 54, 339-342.

PIERS, E. V. Manual of Piers-Harris Children's Self-Concept Scale: (The way I feel about myself). Nashville, T. N.: Counselor Recordings and Tests, 1969.

PURKEY, W. W. Self concept and school achievement. New Jersey: Prentice-Hall, 1970.

RECKLESS, W. C., DINITZ, S. The prevention of juvenile delinquency: An experiment. Columbus, Ohio: Ohio State University Press, 1972.

RECKLESS, W. C., DINITZ, S., KAY, B. The self component in potential delinquency. American Sociological Review, 1957, 22, 566-570.

ROGERS, C. M., SMITH, M. D., COLEMAN, J. M. Social comparison in the classroom: The relationship between academic achievement and self-concept. Journal of Educational Psychology, 1978, 70, 50-57.

ROGERS, C. R. Client-centered therapy. Boston: Houghton Mifflin Company, 1951.

ROGERS, C. R. Freedom to learn. Columbus Ohio: Merrill Publishers Company, 1969.

ROSENBERG, M., SIMMONS, R. G. Black and white self-esteem: The urban school child. Washington, D. C.: American Sociological Association, 1973.

SASSENRATH, J. M. Effects of delayed feedback and length of post feedback interval on retention of prose material. Psychology in Schools, 1972, 9, 194-197.

SASSENRATH, J. M. Theory and results on feedback and retention. Journal of Educational Psychology, 1975, 67, 894-899.

SASSENRATH, J. M., YONGE, G. D. Effects of delayed information feedback and feedback cues in learning on delayed retention. Journal of Educational Psychology, 1969, 60, 174-177.

SASSENRATH, J. M., YONGE, G. D., SHRABLE, K. Immediate and delayed feedback on examinations and immediate and delayed retention. California Journal of Educational Research, 1968, 19, 226-231.

SCHEIRER, M. A., KRAUT, R. E. Increasing educational achievement via self concept change. Review of Educational Research, 1979, 49, 131-150.

SECORD, P. F., BACKMAN, C. W. Social Psychology. Tokyo: McGraw-Hill Kogakusha, LTD., 1964, rep. 1976.

SHANNON, C. E., WEAVER, W. The mathematical theory of communication. Illinois: University of Illinois Press, 1949.

SHRAUGER, J. S., ROSENBERG, S. E. Self-esteem and the effects of success and failure feedback on performance. Journal of Personality, 1970, 38, 404-417.

SILVERMAN, I. Self-esteem and differential responsiveness to success and failure. Journal of Abnormal and Social Psychology, 1964, 69, 115-119.

SKINNER, B. F. Science and human behavior. Macmillan, 1953.

SKINNER, B. F. The science of learning and the art of teaching. Harvard Education Review, 1954, 24, 86-97.

SKINNER, B. F. Teaching machines. Science, 1958, 128, 969-977.

SKINNER, B. F. About behaviorism. New York: Knopf, 1974.

SNYG, D., COMBS, A. W. Individual behavior: A new frame of reference for psychology. New York: Harper, 1949, rev. 1959.

SPARTZ, L. R., SASSENRATH, J. M. Retention of reading material as a function of feedback time and testing. California Journal of Educational Research, 1972, 23, 182-187.

STALLINGs, J. Implementation and child effects of teaching practices in Follow Through classrooms. Monographs of the Society for Research in Child development, 1975, 40, (7-8 #163).

STALLINGS, J. A., KASKOWITZ, D. H. Follow Through classroom observation evaluation. 1972-73, Melo Park, California: Stanford Research Institute, 1974.

STEBBINS, L. B., ST.PIERRE, R. G. PROPER, E. C., ANDERSON, R. B., CERVA, T. R. Education as experimentation: A planned variations model, Vol. IV-A, An evaluation of follow through. Cambridge, Mass.: Abt Associates, 1977.

STURGIS, P. T. Verbal retention as a function of the informativeness and delay of informative feedback. Journal of Educational Psychology, 1969, 60, 11-,4.

SURBER, J. R., ANDERSON, R. C. Delay-retention effect innatural classroom settings. Journal of Educational Psychology, 1975, 67, 170-173.

TAIT, K. HARTLEY, J. R., ANDERSON, R. C. Feedback procedures in computer-assisted arithmetic instruction. British Journal of Educational Psychology, 1973, 43, 161-171.

THORNDIKE, E. L. The fundamentals of learning. Teachers College, Columbia University, 1932.

THORNDIKE, E. L. An experimental study of rewards. Teachers College, Columbia University, Publication #580, 1933a.

THORNDIKE, E. L. A theory of the action of the after-effects of a connection upon it. Psychological Review, 1933, 40, 434-89.

TOLMAN, E. C. Purposive behavior in animals. Berkeley: University of California Press, 1932.

TROWBRIDGE, N. Self-concept and socio-economic status in elementary school children. American Educational Research Journal, 1972, 9, 525-537.

WHITE, K. Delay of test information feedback and learning in a conventional classroom. Psychology in Schools, 1968, 5, 78-81.

WILKS, S. S. Sample criteria for testing equality of menas, equality of variances, and equality of covariances in a normal multivariate distribution. The Annals of Mathematical Statistics, 1946, 17, 257-281.

von WRIGHT, J. M. A note on the role of guidance in learning. British Journal of Psychology, 1957, 48, 133-137.

GULLIKSEN, Harold. Theory of mental tests. New York: John Wiley & Sons, Inc. , 1950.

APPENDIX A.

Domain of the Arithmetic Achievement Test

The criteria which determined the suitability of content for this domain were: i) the type of content (operations in arithmetic) the researcher was interested in; ii) the level and other features of arithmetic content included in the Grade 3 curriculum guide for teachers in Alberta.

The following examples were used to convey the meanings of abreviations employed in this domain definition:

H(C) - canonical horizontal arrangement.
In instances where an H(C) arrangement was used, the problem was presented to the student as follows:

$28 + 7 = \_\_\_$ $28 - 7 = \_\_\_$ $28 \times 7 = \_\_\_$ $28 \div 7 = \_\_\_$;

H(Nc) - noncanonical horizontal arrangement.
In instances where H(Nc) arrangements were chosen, the problem was presented to the student as follows:

$7 + \_\_\_ = 28$ $28 - \_\_\_ = 7$ $7 \times \_\_\_ = 28$ $28 \div \_\_\_ = 7$

$\_\_\_ + 7 = 28$ $\_\_\_ - 7 = 28$ $\_\_\_ \times 7 = 28$ $\_\_\_ \div 7 = 4$

V - vertical arrangement.
Examples of vertical arrangements are as follows:

28	2828	28	2828	77	28	28
+7	+7777	- 7	- 777	-28	X 7	÷ 7
___	____	____	____	____	___	____

c - complete, requiring the filling in of a blank space in a sequence.

r - rewrite, requiring a transformation or an expansion.

n - next, requiring the next higher or the next smaller number.

i - identify, requiring an identification of the correct choice from among several options.

Op.	Format	Content	Features	Replacement scheme	Weight
+	H(C,Nc)	Sum 31 to 49	With re-grouping	3 addends of 1 or 2 digits.	4
-	H(C,Nc)	Minuend 29 to 39	With re-grouping	2 digit sudtra-hends.	2
-	H(Nc)	Minuend 10 to 19	With re-grouping.	1 or two digit subtrahends.	1
+	V	Sum 101 to 999, excl. mult. of 10	With regrouping.	3 addends of 2 or 3 digits.	4
-	V	Minuend 101 to 999.	With re-grouping.	2 or 3 digit subtrahends.	4
-	H(C)	Minuend = 1000	With re-grouping.	one 2 digit sub-trahend between 11 and 14.	1
X	H(C,Nc)	Number facts product 2 to 48.	Excl. mult. of 10.	1 digit factors.	3
X	V	Product 101 to 299	Excl. mult. of 10.	1 digit multiplier.	1
+	V	Sum 1001 to 1100	With re-grouping	2 addends of 3 digits.	1
÷	H(Nc)	Number facts	Dividends to 18	1 digit factors	1
c	H	Counting by 5 and by 10	multiples, before or after	up to the 50th multiple.	1
c	H	Counting by 3 and by 4	multiples, before or after	10th to the 20th multiple.	1
r		Numerals 101 to 999	Excl. mult. of 10.	expansion or trans-formation, e.g. from figures to words.	1
i	pictorial	Fractions 1/2, 1/3, etc. to 1/9		circular, triangular, or rectangular representations.	1

APPENDIX B

ARITHMETIC DRILL

FORM ONE

SCHOOL _____ GRADE _____

NAME _____

BOY _____ OR GIRL _____ BIRTH DATE _____

TODAY'S DATE _____

Cuthbert H. Wright
1930

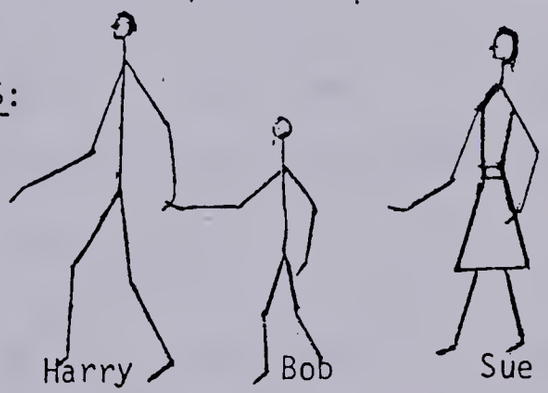
A TEST OF ARITHMETIC OPERATIONS

Directions: This test gives you a chance to practice some of the mathematical skills you have learnt already.

The test contains exercises. Four answers are given for each exercise. You are to choose the ONE answer you think is correct. Then, find the row on the ANSWER SHEET that is numbered the same as the exercise. Now, shade the answer space which matches the answer you have chosen.

Here are two sample exercises. The first one is answered for you.

SAMPLE ITEMS:



1. Who is Bob?

   a) a big man b) a boy

   c) a girl d) a woman

The correct answer is 'a boy'; that is, the choice 'b'. We shade 'b' in the first row to show this.

2. X  Y 

   'X' has more marbles than 'Y'.

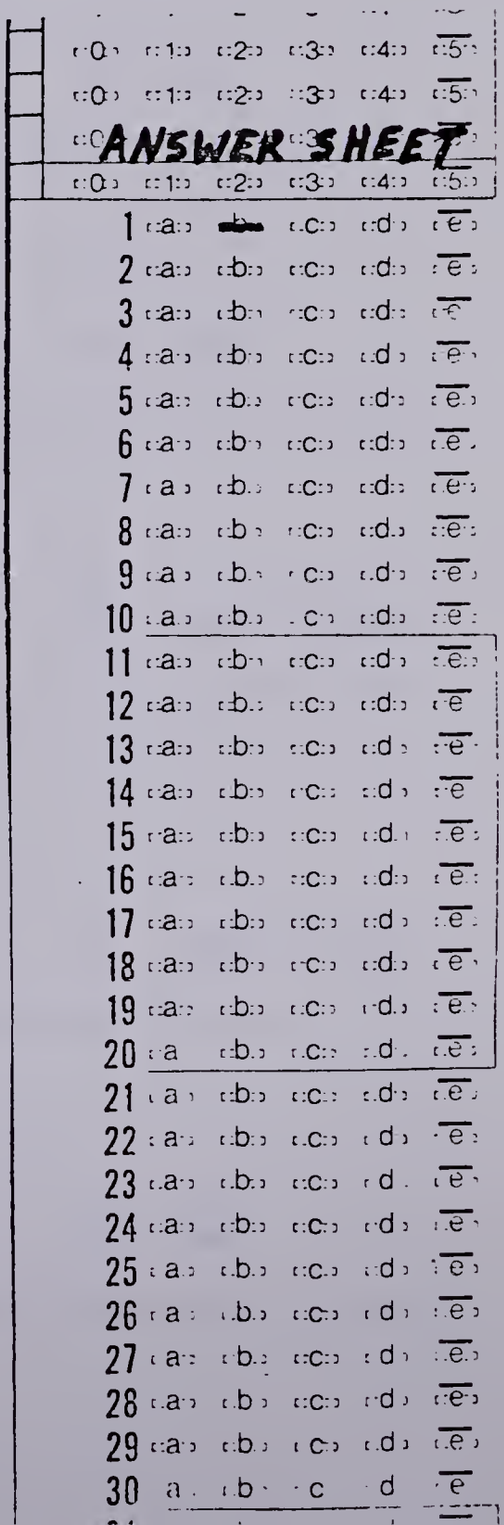
   a) false b) true

   c) both true and false

   d) neither true nor false

The correct answer is 'a'. Show this by finding row 2 (to match exercise 2) and shade the choice 'a' on the ANSWER SHEET pictured on this page. DO IT NOW.

TURN OVER AND BEGIN THE TEST.



1. ____ + 24 + 14 = 48

a) 86 b) 20 c) 10 d) 9

---

2. 17 + _____ + 6 = 32

a) 9 b) 19 c) 11 d) 55

---

3. 19 + 4 + _____ = 37

a) 60 b) 24 c) 14 d) 13

---

4. 36 - 17 = _____

a) 18 b) 19 c) 29 d) 53

---

5. 27 - _____ = 15

a) 42 b) 13 c) 12 d) none of these

---

6. 345
-67

a) 412 b) 378 c) 278 d) none of these

---

7. 38
26
+42

a) 96 b) 106 c) 94 d) 104

---

8. 231
387
+135

a) 643 b) 653 c) 743 d) 753

---

9. 57
158
+369

a) 584 b) 464 c) 474 d) 473

---

10. 37
94
+828

a) 848 b) 859 c) 949 d) 959

---

11. 85
+----
218

a) 293 b) 303 c) 133 d) 143

---

12. 32
X 8

a) 256 b) 4 c) 40 d) 246

---

13. ____ - 14 = 3

a) 143 b) 17 c) 12 d) 11

---

14. 710
-298

a) 512 b) 500 c) 412 d) 426

---

15. 784
-28

a) 766 b) 756 c) 812 d) 752

---

16. _____ X 5 = 5

a) 0 b) 1 c) 2 d) 5

---

17. 5 X _____ = 45

a) 5 b) 8 c) 7 d) 9

---

18. 6 X 4 = ____

a) 64 b) 18 c) 24 d) 28

---

19. _____ ÷ 4 = 2

a) 2 b) 8 c) 6 d) 42

---

20. 585
+418

a) 1003 b) 903 c) 992 d) 993

---

21. 1000 - 11 = ____

a) 989 b) 999 c) 1089 d) 889

---

22. Which pair of numbers should go in the blank spaces?
____, ____, 170, 180, 190,

a) 160, 165, b) 168, 169, c) 160, 161, d) 150, 160,

---

23. Which number best fits the blank space?
____, 40, 44, 48,

a) 41 b) 39 c) 38 d) 36

---

24. The number 142 is equal to which of the following?

a) 1 + 4 + 2 b) 10 + 40 + 2

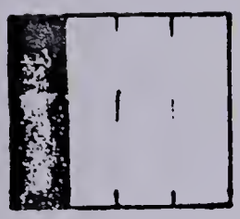
c) 100 + 40 + 2 d) 100 + 4 + 20

---

25. ____ + 8 + 7 = 37

a) 23 b) 22 c) 12 d) 52

26.



1

2

3

4

In which of these figures is ONE-QUARTER shaded?

a) 1 b) 2 c) 3 d) 4

ARITHMETIC DRILL

FORM TWO

SCHOOL _____ GRADE _____

NAME _____

BOY _____ OR GIRL _____ BIRTH DATE _____

TODAY'S DATE _____

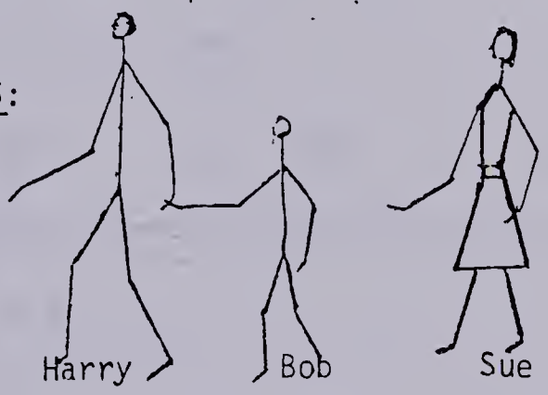
A TEST OF ARITHMETIC OPERATIONS

Directions: This test gives you a chance to practice some of the mathematical skills you have learnt already.

The test contains exercises. Four answers are given for each exercise. You are to choose the ONE answer you think is correct. Then, find the row on the ANSWER SHEET that is numbered the same as the exercise. Now, shade the answer space which matches the answer you have chosen.

Here are two sample exercises. The first one is answered for you.

SAMPLE ITEMS:



1. Who is Bob?

   a) a big man b) a boy

   c) a girl d) a woman

The correct answer is 'a boy'; that is, the choice 'b'. We shade 'b' in the first row to show this.

2. X  Y 

   'X' has more marbles than 'Y'.

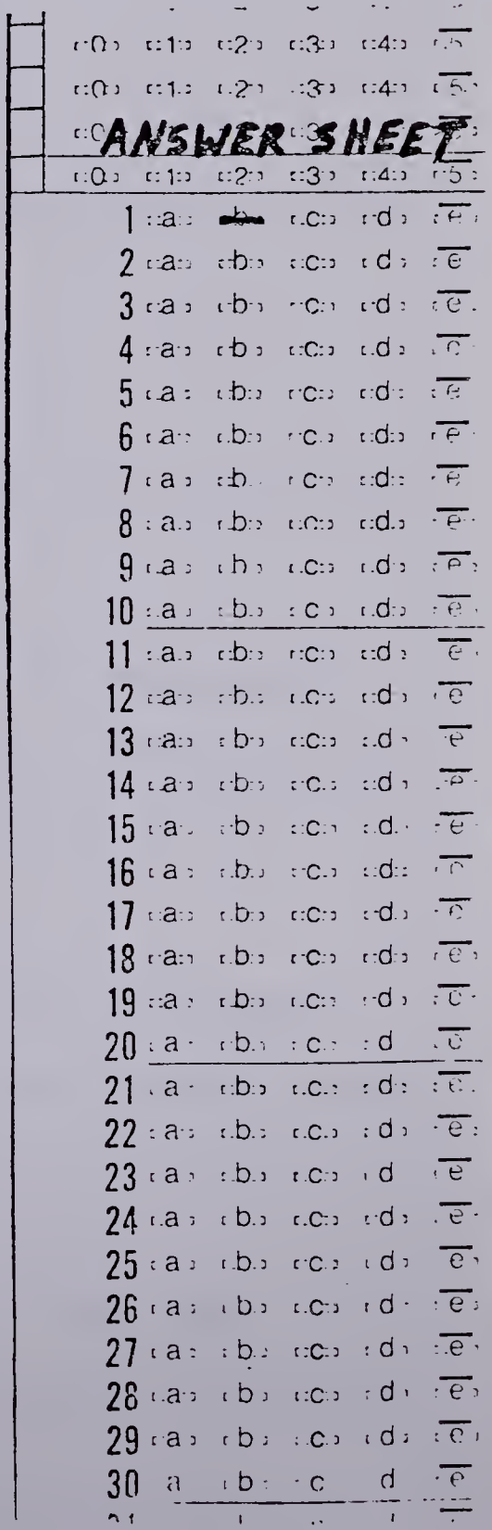
   a) false b) true

   c) both true and false

   d) neither true nor false

The correct answer is 'a'. Show this by finding row 2 (to match exercise 2) and shade the choice 'a' on the ANSWER SHEET pictured on this page. DO IT NOW.

TURN OVER AND BEGIN THE TEST.



1. _____ + 29 + 11 = 47

a) 87 b) 17 c) 8 d) 7

---

2. 14 + _____ + 3 = 41

a) 23 b) 24 c) 34 d) 59

---

3. 17 + 8 + ____ = 44

a) 29 b) 18 c) 19 d) 69

---

4. 35 - 16 = _____

a) 18 b) 19 c) 29 d) 51

---

5. 29 - _____ = 15

a) 44 b) 16 c) 14 d) none of these

---

6. 256
-78

a) 334 b) 278 c) 178 d) none of these

---

7. 27
24
+54

a) 95 b) 105 c) 94 d) 104

8.
$$\begin{array}{r} 159 \\ 419 \\ +184 \\ \hline \end{array}$$

a) 662 b) 762 c) 642 d) 752

9.
$$\begin{array}{r} 58 \\ 279 \\ +137 \\ \hline \end{array}$$

a) 463 b) 374 c) 464 d) 474

10.
$$\begin{array}{r} 77 \\ 46 \\ +326 \\ \hline \end{array}$$

a) 349 b) 339 c) 449 d) 439

11.
$$\begin{array}{r} 51 \\ +--- \\ \hline 343 \end{array}$$

a) 192 b) 292 c) 392 d) 394

12.
$$\begin{array}{r} 24 \\ \times\ 7 \\ \hline \end{array}$$

a) 168 b) 158 c) 33 d) 148

---

13. ____ - 7 = 6

a) 76 b) 1 c) 13 d) 12

---

14. 750
-64

a) 786 b) 696 c) 690 d) 686

---

15. 907
-821

a) 186 b) 106 c) 86 d) 1728

---

16. 9 X ____ = 18

a) 27 b) 2 c) 3 d) 4

---

17. ____ X 6 = 48

a) 6 b) 7 c) 8 d) 9

---

18. 8 X 2 = _____

a) 16 b) 14 c) 24 d) 10

---

19. ____ ÷ 1 = 6

a) 16 b) 7 c) 6 d) 5

---

20. 429
+687

a) 1016 b) 1006 c) 1115 d) 1116

---

21. 1000 - 13 = _____

a) 887 b) 988 c) 997 d) 987

---

22. Which set of numbers should go in the blank spaces?
____, ____, 350, 360, 370,

a) 340, 345, b) 348, 349, c) 330, 340, d) 340, 341,

---

23. Which number best fits the blank space?
_____, 64, 68, 72,

a) 60 b) 62 c) 63 d) 73

---

24. The number 565 is equal to which of the following?

a) 5 + 6 + 5 b) 50 + 60 + 5

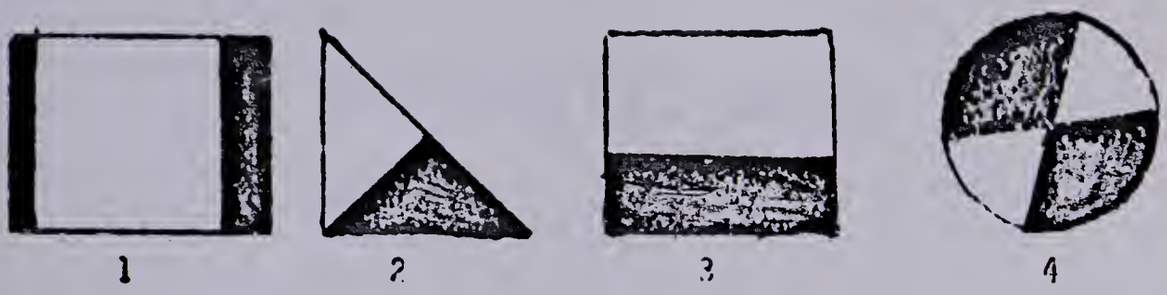
c) 500 + 60 + 5 d) 500 + 6 + 50

---

25. _____ + 4 + 7 = 38

a) 17 b) 29 c) 27 d) 19

26.



1 2 3 4

In which of these figures is ONE-HALF shaded?

a) 1 b) 2 c) 3 d) 4

ARITHMETIC DRILL

FORM THREE

SCHOOL _____ GRADE _____

NAME _____

BOY _____ OR GIRL _____ BIRTH DATE _____

TODAY'S DATE _____

Cuthbert H. Joseph
1950

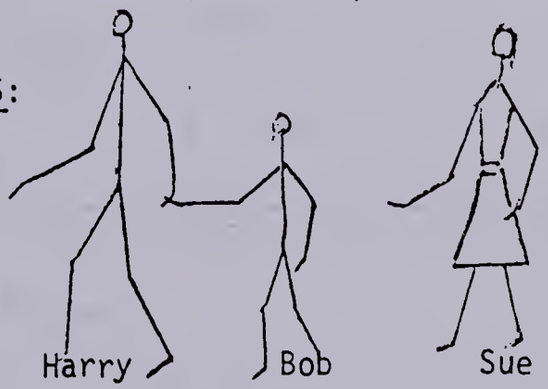
A TEST OF ARITHMETIC OPERATIONS

Directions: This test gives you a chance to practice some of the mathematical skills you have learnt already.

The test contains 26 exercises. Four answers are given for each exercise. You are to choose the ONE answer you think is correct. Then, find the row on the ANSWER SHEET that is numbered the same as the exercise. Now, shade the answer space which matches the answer you have chosen.

Here are two sample exercises. The first one is answered for you.

SAMPLE ITEMS:



1. Who is Bob?

   a) a big man b) a boy
   c) a girl d) a woman

The correct answer is 'a boy'; that is, the choice 'b'. We shade 'b' in the first row to show this.

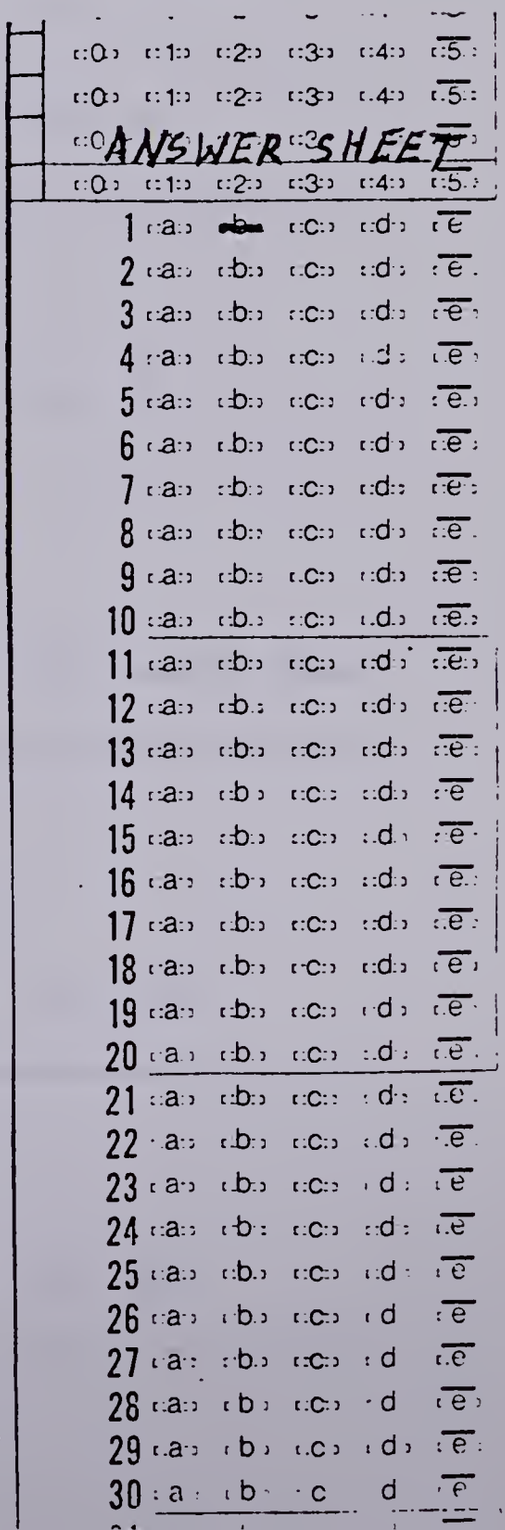
2. X  Y 

   'X' has more marbles than 'Y'.

   a) false b) true
   c) both true and false
   d) neither true nor false

The correct answer is 'a'. Show this by finding row 2 (to match exercise 2) and shade the choice 'a' on the ANSWER SHEET pictured on this page. DO IT NOW.

TURN OVER AND BEGIN THE TEST.



1. _____ + 23 + 17 = 46

a) 86 b) 40 c) 6 d) 7

---

2. 19 + _____ + 7 = 43

a) 27 b) 17 c) 26 d) 69

---

3. 14 + 8 + _____ = 43

a) 21 b) 22 c) 23 d) 65

---

4. 37 - 19 = ____

a) 16 b) 18 c) 17 d) 28

---

5. 29 - _____ = 13

a) 26 b) 17 c) 15 d) none of these

---

6. 24
X 6

a) 144 b) 4 c) 30 d) 124

---

7. 543
-76

a) 719 b) 567 c) 477 d) 467

8.
$$\begin{array}{r} 36 \\ 23 \\ +44 \\ \hline \end{array}$$

a) 93 b) 103 c) 92 d) 102

9.
$$\begin{array}{r} 349 \\ 294 \\ +203 \\ \hline \end{array}$$

a) 736 b) 836 c) 746 d) 846

10.
$$\begin{array}{r} 78 \\ 164 \\ +279 \\ \hline \end{array}$$

a) 421 b) 521 c) 411 d) 511

11.
$$\begin{array}{r} 35 \\ 81 \\ +667 \\ \hline \end{array}$$

a) 673 b) 873 c) 783 d) 773

12.
$$\begin{array}{r} 96 \\ +---- \\ \hline 182 \end{array}$$

a) 278 b) 178 c) 96 d) 86

---

13. _____ - 3 = 7

a) 4 b) 10 c) 37 d) 73

---

14. 340
-53

a) 393 b) 387 c) 287 d) 290

---

15. 961
-484

a) 580 b) 577 c) 477 d) none of these

---

16. _____ X 7 = 28

a) 35 b) 21 c) 6 d) 4

---

17. _____ ÷ 3 = 6

a) 2 b) 9 c) 18 d) 3

---

18. 9 X _____ = 27

a) 3 b) 30 c) 4 d) 5

---

19. 8 X 4 = _____

a) 24 b) 40 c) 16 d) 32

---

20. 432
+589

a) 1021 b) 922 c) 921 d) 911

---

21. 1000 - 12 = _____

a) 998 b) 987 c) 988 d) 1088

---

22. Which pair of numbers should go in the blank spaces?
____, _____, 140, 150, 160,

a) 120, 130, b) 130, 135, c) 130, 131,
d) 138, 139,

---

23. Which number best fits the blank space?
____, 135, 138, 141,

a) 134 b) 133 c) 132 d) 130

---

24. Two hundred fifty-two is equivalent to which of the following?

a) 200502 b) 20052 c) 20502 d) 252

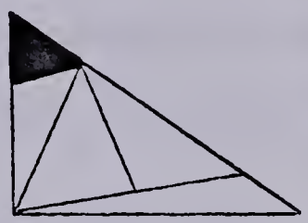
---

25. ____ + 16 + 5 = 29

a) 50 b) 7 c) 8 d) 9

26.

1



2

3

4

In which of these figures is ONE-FIFTH shaded?

a) 1 b) 2 c) 3 d) 4

CURRICULUM VITAE

This curriculum vitae is restricted to my post high school experiences which are most related to the production of this thesis.

September 1965 - 1966 - 1967 - 1968

I taught at an elementary public school for three years starting 1965. As was costumary at Grade 4, this involved teaching all the subjects offered. Consequently, it was here that I was first exposed to the teaching of arithmetic.

September 1968

I was awarded a scholarship to pursue studies at the Teacher Training Department of the Antigua State College. This led to my professional certification as a teacher. Special honours obtained were an A- overall average and an Endorsement awarded by the University of The West Indies.

January 1969

I was elected First Years Representative at the Antigua State College, then The Leeward Islands Teachers Training College.

January 1970

I was elected Headman Student at the Antigua State College, then LITTC.

September 1970

I was transferred as a Certificated Teacher to a Secondary School (high

school). There I taught mathematics at Grades 7, 9 and 11. With more than 2/3 of my weekly teaching schedule (that is, over 20 periods per week) allotted to mathematics, I had to devote much of my time to the close reading of methods, techniques and innovations of mathematics teaching.

September 1970 - August 1973

I was selected to work along with a team of mathematics teachers and consultants in a four-year Modern Mathematics Project. This involved putting together a series of texts and workbooks for junior secondary school children. The texts were entitled Joint School Project Mathematics for the Caribbean, later to be recaptioned Modern Mathematics for the Caribbean. I was required to attend vacation seminars, to try out tests and exercises, and to meet with with a coordinator or a consultant at least once a month.

Summer 1973

I was selected to serve as a co-tutor of Art (Painting) in a Teacher Vacation Course sponsored jointly by the Canadian Teachers' Federation and The Antigua Union of Teachers.

September 1973 - June 1975

I was awarded a CIDA scholarship to complete my first degree at the Mona (Jamaica) campus of the University of the West Indies. The requirements of the degree included the conducting of two separate pieces of research, the first of which was merely a learning experience for the second.

The second of these studies was entitled "A Comparative Study of the Self-Concept of Ability in Mathematics of 'A' and 'D' Stream Eighth Grade Students of Kingston, Jamaica." The results obtained form the sample of 338 Junior Secondary School Eighth Graders indicated that 'A' and 'D' stream eighth graders had similar ($p > .05$) academic self-concept/ arithmetic performance relationships, though the difference between the corresponding relationship was significant when higher and lower performers in each ability stream were compared.

September 1975

I returned to teach mathematics at Grades 9,10, and 11, and at the same time functioned as Head of the school's mathematics department. I held this position until December, 1978 when I left to take up studies here in January of the following year.

Summer 1975, 1976, 1977

I was selected as a Co-tutor of Mathematics in teacher training vacation courses sponsored jointly by the Canadian Teachers' Federation and the Antigua Union of Teachers. Sections of these courses took on workshop type sessions where problems and methods of mathematics teaching were discussed

May 1977

I was elected General Secretary of the Antigua Union of Teachers.

July 1978

Selected as a Co-tutor of a Testing and Measurement workshop for Heads of Schools sponsored jointly by the Organ of Canadian Overseas Development (OCOD and the Ministry of Education, Antigua.